Lecture Notes:
Quantum Information Theory

Gisbert Janßen

December 16, 2019

Contents

About these notes

These notes where written as an accompanying document for a quantum information theory master's course held at the ET-IT department at the Munich Technical University, and therefore adresses primarly an audience with classical information theory backround rather than a formal education in quantum mechanics.
These notes (as well as the course) intend to provide access to the topic for the uninitiated. Therefore, the focus lies on careful introduction of the underlying structure avoiding to praise the folklore. A downside of this approach is a loss in pace. The contents of the course in consequence only cover a very small fraction of what is nowadays canonical. The motivated reader is referred to the exercises.
Especially do these notes not intend to be a textbook on the topic nor should it be regarded as a replacement. The reader is pointed to consult one of the excellent textbooks on the field (we here just mention as examples the books of Wilde [3], and Holevo [2].) Next, we give some overview of the topics of the lectures.

Lecture 1. We introduce the basic entities of quantum theory (for finite-degree systems), and comfort ourselves with its structure by recognizing it as a statistical theory.
Lecture 2. Quantum systems with more than one subsystems are considered. We introduce direct sum and tensor product spaces to describe them. Mathematical tools for describing such systems are introduced. We encounter the first severe "nonclassical" features: Purifications, Entanglement, Noncompatibility.
Lecture 3. As a first asymptotical quantum statistical Problem, we discuss the Quantum Hypotheses Testing Problem. We prove a quantum version of Stein's Lemma which acquaints us with the quantum relative entropy.
Lecture 4. Quantum channels, the broades class of allowed transformations of quantum states are introduced as completely positive and trace preserving linear maps. We add two important equivalent representation theorems (Kraus decomposition and Stinespring dilation) to our toolbox.
Lecture 5. Source compression of discrete memoryless quantum sources ins considered. The quantum source compression theorem is proved. We get to know the quantum Fidelity, and entanglement fidelity as meaningful figures of merit. The optimal compression rate is determined in terms of the von Neumann entropy.
Lecture 6. Transmission of classical messages over disrete and memoryless classical quantum and quantum channels is discussed. The Holevo Theorem is proven to determine the classical message transmission capacity. The coding theorem as well as the converse theorem will be derived as implications of Quantum Stein's Lemma.

While it seems, that the quantum Hypotheses testing problem and Quantum Stein's lemma are usually not regarded as part of the canonical arsenal of first-course quantum information theory topics. In this course, Quantum Stein's lemma

plays a central role. On one hand, good, capacity achieving classical message transmission codes for quantum channels can be derived from certain good hypothesis tests. On the other hand, Quantum Stein's lemma allows for a elementary "information-theoretic" proof of the monotonicity of quantum relative entropy under quantum channels - which helps to derive most of the important entropy inequalities in quantum information theory.
The approach to teaching quantum information theory (and teaching it at all) is inspired by I. Bjelaković and R. Siegmund-Schultze's paper [1], which is strongly recommended as additional reference.

1 Statistical theories in a nutshell

In this course, no prior familiarity with quantum theory is assumed. Since our main interest lies in doing information theory for quantum systems, we take a rather short route to get comfortable with the necessary quantum theoretic terminology.
The approach which best fits our needs, is to regard quantum theory as a mathematical theory to describe the results of a certain type of "statistical" experiments. We already implicitly use such a *statistical theory* ("classical probability theory") when considering classical Shannon theory.
The statistical structure of quantum theory arises from the evidence that microscopic objects (e.g. atoms, photons,...) tend to exhibit "random behaviour" in experiments. Measurement outcomes fluctuate. However, the following assumption seems to be justified.

- If a measurement is performed on independent and equally prepared systems in the same condition many times, the relative frequencies

$$f_i := \frac{\text{number of occurencies of the measurement value } i}{\text{number of total measurements}}$$

  stabilize (i.e. converge to "probabilities")

The above assumption characterizes the hard core of a statistical theory. We will see that there are additional properties a statistical theory should have, and of course there is need to specify, what "measurement", "independent", and "equally prepared" means.
It turned out, that the usual "probability theory" used for describing random experiments with coins (and more complex situations) does not suffice to correctly model some experiments with systems like atoms, photons.
In the next section, we introduce framework which formulates the specifications of *statistical theory* - in generality sufficient to describe the classical and quantum theories.

1.1 Statistical theories - A general framework

A real-world statistical experiment is usually divided into two steps

(i) **Preparation:** Setting a preparation "P" to fix the "initial conditions" of the experiment. Examples

- producing a dice, throwing the dice
- filling an urn with a number of balls having each having some letter in $A..Z$, blindly take a ball from the urn
- setting up a certain laser configuartion for single photon production imprinted.

(ii) **Registration:** Setting a registration "R", i.e. rules how observations are made. Here we assume the most basic type of observation - a "yes/no"-measurement: the registration which one of two given alternatives is taking place. E.g.:

- observing if the number showing up on the dice is odd or even
- observing, whether or not the letter A is printed on the ball
- setting up a semipermeable plate in the laser beam and record whether it went through the plate or not.

Usually, in a statistical context, the experiment is performed many times (lets say a large number N of times), the next step is calculating **relative frequencies**. If N_i is the number of registrations of the ith alternative, the relative frequencies are

$$f_1 = \frac{N_1}{N}, \quad \text{and} \quad f_2 = \frac{N_2}{N} = 1 - f_1$$

Having the statistical postulate from the preceding paragraph in mind, we assume, that the relative frequencies stabilize (i.e. converge) in the limit $N \to \infty$.
Assumption: There is a number $p(P,R)$ – the *probability* of registration.
In order to theoretically describe a certain type of experiment, we form equivalence classes, i.e. we say two preparation procedures P_1 and P_2 are equivalent, if they lead to the same probabilities for all registration procedures, i.e. $p(P_1,R) = p(P_2,R)$ for all possible R. Two registration procedures R_1, R_2 are equivalent, if $p(P,R_1) = p(P,R_2)$ for all preparation procedures P. We call each equivalence class of preparation procedures a *state*, and each equivalence class of registration procedures an *effect*.

(iv) **Theoretical description:** A statistical theory is given by a set $\mathfrak{S}$ of states and a set $\mathfrak{E}$ of effects together with a map $p: \mathfrak{S} \times \mathfrak{E} \to [0,1]$, $(s,E) \mapsto p(s,E)$ with

$$p(s_1,E) = p(s_2,E) \text{ for all } E \in \mathfrak{E} \Rightarrow s_1 = s_2$$
$$p(s,E_1) = p(s,E_2) \text{ for all } s \in \mathfrak{S} \Rightarrow E_1 = E_2$$

we denote the *certain* effect by $\mathbb{1}$ (i.e. $\mathbb{1}$ is the unique effect with $p(s,\mathbb{1}) = 1$ for all states)

Before we proceed with our exposition of statistical theories, we give the two examples which will be of interest in this course. The first one describes experiments with a "classical system", while the second, regards a "quantum system".

Example 1 (Classical statistics)**.** *Let Ω be a finite sample set. The statistical theory usually imposed is given by the state set $\mathfrak{S} := \mathcal{P}(\mathcal{X})$, and effect set $\mathfrak{E} := \mathfrak{E}(\Omega)$ with*

$$\mathcal{P}(\Omega) := \left\{ q : \Omega \to [0,1] : \sum_{\omega \in \Omega} q(\omega) = 1 \right\}$$
$$\mathcal{E}(\Omega) := \{ f : \Omega \to [0,1] : 0 \le f(\omega) \le 1 \text{ for all } \omega \in \Omega \}.$$

The probability to register the effect $E \in \mathfrak{E}$ when preparing with $q \in \mathcal{S}$ is

$$p(q,E) := \sum_{\omega \in \Omega} q(\omega) \cdot f(\omega)$$

Note that in the traditional formulation of probability theory after Kolmogorov, rather *events* are considered instead of effects. This approach is recovered by restricting the registration effects to the set

$$\{\mathbb{1}_A : A \subset \Omega\} \subsetneq \mathfrak{E}(\Omega),$$

where $\mathbb{1}_A$ is the indicator function of A, i.e.

$$\mathbb{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

By using the broader concept of effects, we allow also "fuzzy" registrations.

Example 2 (Quantum statistics)**.** *Consider a finite-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^d$, $d < \infty$, the usual statistical model for a quantum system with d degrees of freedom is given by the state set $\mathfrak{S} := \mathcal{S}(\mathcal{H})$, and $\mathfrak{E} := \mathcal{E}(\mathcal{H})$, where*

$$\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{L}(\mathcal{H}) : \rho^* = \rho \wedge \rho \geq 0 \wedge \mathrm{tr}\rho = 1\} \qquad \textit{(density matrices)}$$
$$\mathcal{E}(\mathcal{H}) := \{E \in \mathcal{L}(\mathcal{H}) : 0 \leq E \leq \mathbb{1}\} \qquad \textit{(quantum effects)}$$

Moreover, for given density matrix ρ, and effect E, the probability of registration is calculated via the formula

$$p(\rho, E) = \mathrm{tr}\rho E \qquad \textit{(Born's rule).}$$

We consider some more aspects of the notions we just introduced.
Convexity: When conducting statistical experiments, there usually also is the principal possibility to *mix* preparation procedures as well as registration procedures. For example, when having two preparation devices P_1, P_2 at hand, one could let a random number generator decide which one of these to take for the next sample. If λ is the probability, that P_1 is chosen, it makes sense to demand, that the resulting preparation procedure $\tilde{P}$ is also allowed, and fulfills

$$p(\tilde{P}, E) = \lambda \cdot p(P_1, R) + (1 - \lambda) \cdot p(P_2, R) \tag{1.1}$$

for each registration procedure R. Formulating this equation on the level of states and effects, we have

$$\lambda p(\rho_1, E) + (1 - \lambda) p(\rho_2, E) = \lambda \mathrm{tr}\rho_1 E + (1 - \lambda)\mathrm{tr}\rho_2 E \tag{1.2}$$
$$= \mathrm{tr}(\lambda\rho_1 + (1 - \lambda)\rho_2)E, \tag{1.3}$$

i.e. the state corresponding to a mixture of states ρ_1, ρ_2 with mixing parameter λ is the corresponding convex combination $\lambda\rho_1 + (1 - \lambda)\rho_2$ for each effect E.
A similar argument can be drawn for mixtures of effects. In consequence, the sets $\mathcal{S}(\mathcal{H})$ and $\mathcal{E}(\mathcal{H})$ are convex subsets of a linear space! It is of interests to know the extremal elements of a convex set (i.e. the elements which are not nontrivial convex combinations of other elements of that set). The extremal elements of the set $\mathfrak{S}$ are called *pure states* (accordingly, states which are not pure are called *mixed*), while the extremal elements of $\mathfrak{E}$ are called *propositions*. In case of quantum theory, we have

Proposition 3. *Let $\mathcal{H}$ be (finite dimensional) Hilbert space. The following claims hold.*

1. *$\rho \in \mathcal{S}(\mathcal{H})$ is a pure state if and only if it is a rank one projection.*
2. *$E \in \mathcal{E}(\mathcal{H})$ is a proposition if and only if it is a projection.*

Proof. We show the first claim. The second can be proven by similar arguments. We first show the "$\Rightarrow$" implication. Assume, that ρ is a rank one projection, i.e. it holds $\rho^2 = \rho$. We show, that any convex combination

$$\rho = \mu\tau_1 + (1-\mu)\tau_2$$

is necessarily trivial. It holds

$$\begin{aligned}\rho - \rho^2 &= \mu\tau_1 + (1-\mu)\tau_2 - \mu^2\tau_1^2 - \mu(1-\mu)(\tau_1\tau_2 + \tau_2\tau_1) - (1-\mu)^2\tau_2^2 \\ &= \mu(\tau_1 - \tau_1^2) + (1-\mu)(\tau_2 - \tau_2^2) + \mu(1-\mu)(\tau_1 - \tau_2)^2 \\ &\geq \mu(1-\mu)^2(\tau_1 - \tau_2)^2 \\ &\geq 0.\end{aligned}$$

The equalities above are by rearranging terms. The first inequality above (notice: the inequality is a matrix inequality in the hermitian semiorder) is by the fact, that μ and $1-\mu$ are nonnegative and $(\tau_1 - \tau_1^2)$ as well as $\tau_2 - \tau_2^2$ are positive semidefinite matrices (check this.) The second inequality follows, because $(\tau_1 - \tau_2)^2$ is positive semidefinite. Since $\rho - \rho^2 = 0$ (ρ is assumed to be a projections), it holds

$$\mu(1-\mu)(\tau_1 - \tau_2)^2 = 0. \tag{1.4}$$

But this is only possible if $\mu \in \{0, 1\}$ or $\tau_1 = \tau_2$, which is the case if and only if the convex combination in Eq. (1.4) is trivial.
For the showing remaining "$\Leftarrow$" implication, let ρ be an extremal element of $\mathcal{S}(\mathcal{H})$. Consider a spectral decomposition

$$\rho = \sum_{i=1}^{\dim \mathcal{H}} \lambda_i |\psi_i\rangle\langle\psi_i| \tag{1.5}$$

of ρ. Notice that this is a convex combination of ρ. Since ρ is assumed to be extremal there is exactly one i_0 with $\lambda_{i_0} = 1$, while $\lambda_i = 0$ for all $i \neq i_0$. Consequently $\rho = |\psi_{i_0}\rangle\langle\psi_{i_0}|$, a rank one projection. □

Remark 4. *Each unit vector $v \in \mathcal{H}$ gives rise to a pure state $\rho = |v\rangle\langle v|$. the correspondence $v \leftrightarrow |v\rangle\langle v|$ is one-to-one up to global phases, i.e. for $\theta \in \mathbb{R}$, v and $e^{i\theta}v$ give rise to the same pure state.*

To describe statistical experiments with more than two outcomes, we introduce the concept of an observable. An *observable* (or *measurement*) with a (finite) set $\mathcal{Y}$ of measurement outcomes is a function $F : \mathcal{Y} \to \mathfrak{E}$ such that

$$\sum_{y \in \mathcal{Y}} E_y \;=\; \mathbb{1} \tag{1.6}$$

Since the set of measurement values is finite, it is more common to define a measurement by a collection of effects. In case of quantum theory, we define

Definition 5. *Let $\mathcal{H}$ be a finite dimensional Hilbert space, and $\mathcal{Y}$ be a finite set. A POVM (positive operator valued measure) on $\mathcal{H}$ with measurement outcomes in $\mathcal{Y}$ is a family $\{E_y\}_{y\in\mathcal{Y}}$ such that*

1. $0 \leq E_y \leq \mathbb{1}_{\mathcal{H}}$ *for all* $y \in \mathcal{Y}$
2. $\sum_{y\in\mathcal{Y}} E_y = \mathbb{1}_{\mathcal{H}}$.

The special case, of a family of mutually orthogonal projections in $\mathcal{H}$ is called projection valued measure (PVM) *or simply* von Neumann measurement .

In this course we restrict ourselves to finite dimensional Hilbert spaces and finite sets of measurement values, which avoids topological and measure theoretic issues. The interested reader may consult the following references for more details.

1. K. Kraus, *States, Effects and Operations*, Springer, 1983
2. A. Holevo, *Quantum Systems, Channels, Information*, de Gruyter, 2012, Chapter 2.
3. A. Holevo, *Probabilistic and Statistical Aspcects of Quantum Theory*, Edizione de Scuola Normale Superiore Pisa, 2011, Chaper 1
4. S. Gudder, *Stochastic Methods of Quantum Mechanics*, Dover Publications, 2005, Chapter 4

1.2 Example: Qubit systems

To get into calculations with the mathematical objects defined above, we consider the case of a quantum system with two degrees of freedom, i.e. the underlying Hilbert space is two-dimensional. Despite the fact, that such systems are often considered in physics[1], they can be regarded as quantum counterparts of classical bit systems bit systems with alphabet $|\mathcal{X}| = 2$. The set $\mathcal{S}(\mathbb{C}^2)$ of qubit states has a convenient pictoral representation in $\mathbb{R}^3$, which we derive next. The matrices

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are called *Pauli matrices* and form an orthogonal basis in $\mathcal{L}(\mathbb{C}^2)$, it holds

$$\langle \sigma_i, \sigma_j \rangle_{HS} = 2\delta_{ij}. \tag{1.7}$$

If we normalize each of the matrices with a factor $1/\sqrt{2}$ we obtain an orthonormal basis. We can write each matrix $A \in \mathcal{L}(\mathcal{H})$ as a linear combination

$$A = \frac{1}{2}\sum_{i=0}^{3} r_i \sigma_i. \tag{1.8}$$

[1] Real-world examples of such systems are e.g. the spin of an electron system or the polarization of light.

with $r_i = \langle A, \sigma_i \rangle_{HS}$. We aim to derive conditions on the numbers $r_0, \ldots, r_3$ being equivalent to $A \in \mathcal{S}(\mathcal{H})$. We have

1. Each r_i has to be real, because A is Hermitian.
2. $r_0 = 1$ holds because of the property $\mathrm{tr}(A) = 1$.
3. Since $A \geq 0$ holds,

$$\frac{1}{4}(r_0^2 - r_1^2 - r_2^2 - r_3^2) = \det A \geq 0. \tag{1.9}$$

Since $r_0 = 1$, we obtain the condition $\|r\| \leq 1$ for the vector $(r_1, r_2, r_3)^T \in \mathbb{R}^3$.

On the other hand, if A is represented as in (1.8) with $r_0 =$ and $(r_1, r_2, r_3)^T$ an element of $B_1(0)$, the euclidean ball around 0 with radius one, then $A \in \mathcal{S}(\mathcal{H})$ is implied. Indeed, $1 = r_0 = \mathrm{tr}A$, and $\det A = 1 - r_1^2 - r_2^2 - r_3^2 \geq 0$. Consequently, A is a density matrix. Since the basis coefficients r_0, r_1, r_2, r_3 of an element of $\mathcal{L}(\mathcal{H})$ are unique, the map which connects each density matrix with its bloch vector $(r_1, r_2, r_3)^T$ is a one-to-one. By linearity of the Hilbert-Schmidt scalar product, it is clear that this map is also affine. Therefore, we have introduced an affine bijection of $\mathcal{S}(\mathcal{H})$ onto the radius one euclidean ball in $\mathbb{R}^3$. By this fact, it is clear, that the set of extremals of $\mathcal{S}(\mathcal{H})$ correspond to the set of extremals of $B_1(0)$, i.e. the unit sphere around 0 in $\mathbb{R}^3$.

Remark 6. *In quantum optics, it is common, to specify the polarization preparation of a leaser beam by giving the corresponding bloch vector $r = (r_1, r_2, r_3)^T$. Examples are (according to www.wikipedia.de)*

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \textit{ (linear horizontal)} \qquad \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \textit{ (linear vertical)}$$

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \textit{ (linear 45°)} \qquad \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \textit{ (linear -45°)}$$

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \textit{ (right circular)} \qquad \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \textit{ (left circular)}$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \textit{ (unpolarized).}$$

By calculating their bloch vectors, one can verify, that the pure states $P_i := |e_i\rangle\langle e_i|$, $i \in \{0, 1\}$ are located at the north and south poles of the Bloch ball. The set of states which lie on the straight line connecting the pure states P_0 and P_1 are parameterized by probability distributions on $\{0, 1\}$,

$$\{P(p) := p(0)P_0 + p(1)P_1\}.$$

In fact, each straight line connecting two antipodes on the Bloch sphere can be regarded as a version of the classical bit states.

2 Composite quantum systems

In the preceding lecture, we introduced the basic quantum-theoretic entities for given quantum system with Hilbert space $\mathcal{H}$. In this lecture, we consider quantum systems which are composed of two or more *subsystems*. E.g. we need to be able to describe the statistics of experiments where two electron spins (each with Hilbert space $\mathcal{H} = \mathbb{C}^2$) are involved. It is one of the major tasks in information theory, to quantify correlations between such subsystems and optimally exploit statistical properties of composite systems for usage in communication protocols.
Remembering classical statistical theory (with finite sample spaces), the state of a system composed of $n < \infty$ subsystems with alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_n$ the set of states on that system is the set $\mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$ of probability distributions on the cartesian product of $\mathcal{X}_1, \ldots, \mathcal{X}_n$. We notice, that the pure states are exactly the products of point measures, $\{\delta_{x^n} : x^n \in \mathcal{X}^n\}$ with

$$\delta_{x^n}(y^n) := \prod_{i=1}^{n} \delta_{x_i}(y_i) \qquad (x^n = (x_1, \ldots, x_n), y^n = (y_1, \ldots, y_n)).$$

Taking as an example $n = 2$, we can define to each $p \in \mathcal{P}(\mathcal{X} \times \mathcal{X}_2)$ the *marginal states* $p_1 \in \mathcal{P}(\mathcal{X}_1), p_2 \in \mathcal{P}(\mathcal{X}_2)$ by

$$p_1(x_1) := \sum_{x_2 \in \mathcal{X}_2} p(x_1, x_2), \quad \text{and} \qquad p_2(x_2) := \sum_{x_1 \in \mathcal{X}_1} p(x_1, x_2)$$

which recover the statistical properties of the individual subsystems. The systems are statistically *independent*, if the state p is a product distribution, i.e. with some $p_1 \in \mathcal{P}(\mathcal{X}_1),\ p_2 \in \mathcal{P}(\mathcal{X}_2)$ it holds

$$p(A \times B) \ = \ p_1(A) \cdot p_2(B)$$

for all $A \subset \mathcal{X}_1$, $B \subset \mathcal{X}_2$ (the usual notation is then "$p = p_1 \otimes p_2$".) We notice, that each $w \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ can be written as a convex combination of product states via

$$w \ = \ \sum_{x^n \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n} w(x^n)\ \delta_{x^n}.$$

Regarding a composite quantum system with subsystem Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_n$, we need to define the states on that Hilbert space. Therefore, the cartesian product $\mathcal{H}_1 \times \cdots \times \mathcal{H}_N$ has to be equipped with a Hilbert space structure.

Mathematical Interlude – Direct sums and tensor products

In this course, we will make extensive use of two different possibilites to provide the cartesian product

$$V_1 \times \cdots \times V_N \ := \ \{(v_1, \ldots, v_N) : \ v_i \in V_i,\ 1 \le i \le N\}$$

of linear spaces $V_1, \ldots, V_n$ with a linear structure, the *direct sum* $V_1 \oplus \cdots \oplus V_n$ and the *tensor product* $V_1 \otimes \cdots \otimes V_n$. Since we are always dealing with Euclidean spaces, we assume V_i to be equipped with a scalar product $\langle \cdot, \cdot \rangle_i$, $i = 1, \ldots, n$.

Direct sum spaces

The direct sum $V_1 \oplus \cdots \oplus V_N$ of $V_1, \ldots, V_N$ is defined as the set

$$\left\{ \begin{pmatrix} v \\ w \end{pmatrix} : \; v \in V_1, \; w \in V_2 \right\}$$

together with the obvious linear structure inherited from the component spaces by using addition and scalar multiplication component-wise, i.e.

$$\lambda \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} = \begin{pmatrix} \lambda \cdot v_1 \\ \vdots \\ \lambda \cdot v_N \end{pmatrix} \qquad (\lambda \in \mathbb{C}, v_1 \in V_i, \; 1 \leq i \leq N), \text{ and}$$

$$\begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} + \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_N + w_N \end{pmatrix} \qquad (v_i, w_i \in V_i, \; 1 \leq i \leq N).$$

We may also define a scalar product $\langle \cdot, \cdot \rangle_{1 \oplus \cdots \oplus N} : V_1 \cdots \oplus V_N \to \mathbb{C}$ by setting

$$\left\langle \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} \right\rangle_{1 \oplus \cdots \oplus N} := \sum_{i=1}^{N} \langle v_i, w_i \rangle_i$$

for all $v_i, w_i \in V_i$, $1 \leq i \leq N$. On the other hand, if $\mathcal{H}$ is a Hilbert space, and $V_1, \ldots, V_N \subset \mathcal{H}$ are pairwise orthogonal linear subspaces such that for each $h \in \mathcal{H}$ exist uniquely $h_1 \in V_1, \ldots, h_N \in V_N$ such that $h = h_1 + \cdots + h_N$, then $\mathcal{H}$ is isomorphic to $V_1 \oplus \cdots \oplus V_N$. We can define injections

$$I_j \in \mathcal{L}(V_j, \mathcal{H}), \;\; v \mapsto I_j(v) = (0, \ldots, v, \ldots, 0) \qquad (v \in V_j).$$

Notice, that the adjoints are given by $I_j^*(h) = P_j h$, where P_j is the projector onto V_j for $j \in [N]$. Choosing an appropriate orthonormal basis, each $A \in \mathcal{L}(\mathcal{H})$ can be written as a *block matrix*

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{pmatrix}$$

where $A_{ij} = I_j^* A I_i \in \mathcal{L}(V_i, V_j)$ is a smaller matrix for all $i, j \in [N]$.

Exercise 7. *Show, that if* $\{v_i\}_{i=1}^{\dim V_1} \subset V_1$ *and* $\{w_j\}_{j=1}^{\dim V_2} \subset V_2$ *are orthonormal bases in* V_1, V_2, *then*

$$\left\{ \begin{pmatrix} v_i \\ 0 \end{pmatrix} \right\}_{i=1}^{\dim V_1} \cup \left\{ \begin{pmatrix} 0 \\ w_i \end{pmatrix} \right\}_{j=1}^{\dim V_2} \tag{2.1}$$

is an orthonormal basis in $V_1 \oplus V_2$ *according to the corresponding scalar product. In particular,* $\dim V_1 \oplus V_2 = \dim V_1 + \dim V_2$.

Tensor product spaces

Another way to equip the cartesian product of $V_1,\dots,V_N$ with a linear structure is forming the tensor product. First, we note, that each pair $(v_1,\dots,v_N)\in V_1\times\cdots\times V_N$ defines a N-linear function $v_1\otimes\otimes v_N : V_1\times\cdots\times V_N\to\mathbb{C}$ by

$$v_1\otimes\cdots\otimes v_N(x_1,\dots,x_N) := \langle v_1,x_1\rangle_1\cdot\dots\cdot\langle v_N,x_N\rangle_N\,.$$

The definitions can be obviously extended to form linear combinations of these *elementary tensors*, i.e. the rules

$$\begin{aligned}
\lambda(v\otimes w) = (\lambda v)\otimes w = v\otimes(\lambda w) \qquad & (\lambda\in\mathbb{C}, v\in V_1, w\in V_2),\\
(v_1+v_2)\otimes w = v_1\otimes w + v_2\otimes w \qquad & (v_1,v_2\in V_1, w\in V_2)\\
v\otimes(w_1+w_2) = v\otimes w_1 + v\otimes w_2 \qquad & (v\in V_1, w_1,w_2\in V_2)
\end{aligned}$$

apply (the rules above are formulated for $N=2$, but it is quite clear how the version for general finite N looks). We can form formal linear combinations

$$\sum_{i=1}^{N}\alpha_{i_1\dots i_N}v_{i_1}\otimes\cdots\otimes v_{i_N}$$

of elementary tensor product vectors with coefficients $\alpha_{i_1\dots i_N}\in\mathbb{C}$. The set

$$V_1\otimes\cdots\otimes V_N := \operatorname{span}\{v_1\otimes\cdots\otimes v_N :\ v_i\in V_i,\ 1\le i\le N\}$$

is called the tensor product of V_1 and V_2. We also can extend these structures to the linear maps. With spaces $V_1,\dots,V_N,\tilde V_1,\dots\tilde V_N$, we define the tensor product $\mathcal{L}(V_1,\tilde V_1)\otimes\cdots\otimes\mathcal{L}(V_N,\tilde V_N)$ accordingly. For each $A_1,\dots,A_N, A_i\in\mathcal{L}(V_i,\tilde V_i)$, $A_1\otimes\cdots\otimes A_N$ is the map defined by

$$(A\otimes\cdots\otimes A_N)(v_1\otimes\cdots\otimes v_N) = A_1v_1\otimes\cdots\otimes A_Nv_N$$

for each $v_1\in V_1,\dots,v_N\in V_N$. We have

$$\begin{aligned}
\mathcal{L}(V_1\otimes V_2,\tilde V_1\otimes\tilde V_2) &= \mathcal{L}(V_1,\tilde V_1)\otimes\cdots\otimes\mathcal{L}(V_N,\tilde V_N)\\
&= \operatorname{span}\{A_1\otimes\cdots A_N :\ A_i\in\mathcal{L}(V_i,\tilde V_i),\ 1\le i\le N\}.
\end{aligned}$$

Isomorphisms and representations

In calculations and proofs, it is sometimes advantageous, to transfer the objects under investigation to another space. In the following we collect some isomorphisms which are often used. The definitions below are defined for members of the canonical euclidean orthonormal basis, but extend – by linearity – to all elements of the underlying spaces.

1.

$$\Sigma :\ \bigoplus_{i=1}^{N}\mathcal{H}\ \to\ \mathcal{H}\otimes\mathbb{C}^N,\qquad \begin{pmatrix} v_1\\ \vdots\\ v_N\end{pmatrix}\mapsto\sum_{i=1}^{N}v_i\otimes e_i$$

2.

$$\Gamma:\ \mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)\ \to\ \mathcal{H}_1\otimes\mathcal{H}_2 \qquad |v\rangle\langle w|\ \mapsto\ |\overline{w}\rangle\otimes|v\rangle \tag{2.2}$$

3.

$$\Lambda:\ \mathcal{L}(\mathbb{C}^M\otimes\mathbb{C}^N)\ \to\ \mathcal{L}\left(\bigoplus_{i=1}^{M}\mathbb{C}^N\right) \qquad |e_i\rangle\langle e_j|\otimes|e_k\rangle\langle e_l|\ \mapsto I_i^*|e_k\rangle\langle e_l|I_j$$

With the preparations, we are able to define quantum states on composite systems.

Definition 8. *The set of states of a composite system with $N<\infty$ subsystems, $\mathcal{H}_i$ being the Hilbert space of the i-th subsystem is given by the set*

$$\mathcal{S}\left(\bigotimes_{i=1}^{N}\mathcal{H}_i\right)$$

of density matrices on the tensor product $\bigotimes_{i=1}^{N}\mathcal{H}_i := \mathcal{H}_1\otimes\cdots\otimes\mathcal{H}_N$ of the spaces $\mathcal{H}_1,\dots\mathcal{H}_N$.

A corresponding map to form "marginal states" on subsystems is given by the partial trace, which we define for notational simplicity for bipartite systems (i.e. composite systems consisting of two subsystems).

Definition 9 (Partial trace)**.** *For two Hilbert spaces $\mathcal{H}_1,\mathcal{H}_2$, the partial trace (over $\mathcal{H}_2$) is the linear map*

$$\mathrm{tr}_{\mathcal{H}_2}:\ \mathcal{L}(\mathcal{H}_1\otimes\mathcal{H}_2)\to\mathcal{L}(\mathcal{H}_1)$$

defined by

$$\mathrm{tr}_{\mathcal{H}_2}(|u\otimes v\rangle\langle w\otimes x|) := \langle x,v\rangle|u\rangle\langle w| \qquad (u,w\in\mathcal{H}_1,\ v,x\in\mathcal{H}_2)$$

plus linear extension. The corresponding partial trace over $\mathcal{H}_1$, $\mathrm{tr}_{\mathcal{H}_1}$ is defined analogeously.

We demonstrate by example, how the partial trace can be calculated. Let $\mathcal{H}_1,\mathcal{H}_2$ be Hilbert spaces, $\dim\mathcal{H}_i=d_i$, $i=1,2$. Let $A\in\mathcal{L}(\mathcal{H}_1\otimes\mathcal{H}_2)$, with orthonormal tensor basis decomposition

$$A=\sum_{i,j=1}^{d_1}\sum_{k,l=1}^{d_2} a_{ijkl}\ |f_i\otimes g_k\rangle\langle f_j\otimes g_l|.$$

Then, we calculate

$$\begin{aligned}\mathrm{tr}_{\mathcal{H}_2}(A) &= \sum_{i,j=1}^{d_1}\sum_{k,l=1}^{d_2} a_{ijkl}\ \mathrm{tr}_{\mathcal{H}_2}(|f_i\otimes g_k\rangle\langle f_j\otimes g_l|)\\ &= \sum_{i,j=1}^{d_1}\sum_{k}^{d_2} a_{ijkl}\ \langle f_l,f_k\rangle\cdot|f_i\rangle\langle f_j|\\ &= \sum_{i,j=1}^{d_1}\left(\sum_{k}^{d_2} a_{ijkk}\right)\cdot|f_i\rangle\langle f_j|.\end{aligned}$$

Note, that the definition of partial traces easily extends to composite systems with more than two subsystems. Let N parties (each of them with Hilbert space $\mathcal{K}_i$ assigned, share a system with Hilbert space $\mathcal{K}_1 \otimes \cdots \otimes \mathcal{K}_N$. To apply the partial trace on the j-th system one uses the definition above with $\mathcal{H}_1 = \bigotimes_{i \neq j} \mathcal{K}_i$, $\mathcal{H}_2 := \mathcal{H}_i$ and so on...

Proposition 10. *Let $A \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. It holds*

1. $\mathrm{tr}(A) = \mathrm{tr}(\mathrm{tr}_{\mathcal{H}_1}(A)) = \mathrm{tr}(\mathrm{tr}_{\mathcal{H}_2}(A))$,
2. $\mathrm{tr}(\mathrm{tr}_{\mathcal{H}_2}(A)B) = \mathrm{tr}(A(B \otimes \mathbb{1}_{\mathcal{H}_2}))$,
3. $A \geq 0 \;\Rightarrow\; \mathrm{tr}_{\mathcal{H}_2}(A) \geq 0$,
4. $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2) \;\Rightarrow\; \mathrm{tr}_{\mathcal{H}_2}(\rho) \in \mathcal{S}(\mathcal{H}_2)$.

Proof. The claims of the proposition can be verified by straightforward calculation, left as an exercise. As an example, we show the second claim. Let $x \in \mathcal{H}_1$, $\|x\| = 1$. Note, that

$$P := |x\rangle\langle x| \otimes \mathbb{1}_{\mathcal{H}_2}$$

is a projection. We have

$$\langle x, \mathrm{tr}_{\mathcal{H}_2}(A)x\rangle \;=\; \mathrm{tr}(|x\rangle\langle x|\mathrm{tr}_{\mathcal{H}_2}(A)) \;=\; \mathrm{tr}(A^{\frac{1}{2}}PA^{\frac{1}{2}}) \;\geq\; 0.$$

□

Here, we may point out a significant difference between the concept of a bipartite state in classical theory and quantum theory. As noticed earlier, each classical bipartite state (i.e. probability distribution on a product alphabet) can be written as a convex combination of product distributions. Quantum probability offers an additional class of states beyond.

Definition 11. *A state $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is called*

(i) "uncorrelated", "product state" *, if it can be written $\rho = \rho_1 \otimes \rho_2$ for some $\rho_1 \in \mathcal{S}(\mathcal{H}_1)$, $\rho_2 \in \mathcal{S}(\mathcal{H}_2)$.*

(ii) "separable state" *, if it admits the form*

$$\rho = \sum_{i=1}^{N} \lambda_i \, \rho_1^{(i)} \otimes \rho_2^{(i)}$$

with $N \in \mathbb{N}$, $\lambda_i \in (0,1)$, $\rho_1^{(i)} \in \mathcal{S}(\mathcal{H}_1)$, and $\rho_2^{(i)} \in \mathcal{S}(\mathcal{H}_2)$ for all $i \in [N]$, $\sum_{i=1}^{N} \lambda_i = 1$.

(iii) entangled *otherwise.*

Remark 12. *One could ask about infinite or even uncountable convex combinations of product states. These are identified as uncorrelated by means of Caratheodory's theorem. It asserts, that for given convex subset $A \subset \mathbb{R}^d$, each element $x \in A$ can be written as a convex combination of at most $d+1$ extremal elements in A. Consequently, each separable state can be written as a finite convex combination of not more than $2d+1$ product states.*

A closer look at Definition 11 reveals, that beyond the convex combinations of product states (of which the product states itself are a trivial subclass), there is another class of states having no classical counterpart. The class of entangled states is indeed nonempty, as the following example demonstrates. Let two unit vectors $\varphi_1, \varphi_2 \in \mathbb{C}^d$ with $\varphi_1 \perp \varphi_2$, and nonzero coefficients $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2 + |\beta|^2 = 1$ be given. We show, that

$$\eta \;=\; \alpha\varphi_1 \otimes \varphi_1 + \beta\varphi_2 \otimes \varphi_2 \tag{2.3}$$

is state vector of an entangled state. In fact, the assumption, that η is separable leads to a contradiction. Let $\{\varphi_1, \varphi_2, \ldots, \varphi_d\}$ be an extension of φ_1, φ_2 to an orthonormal basis in $\mathbb{C}^d$. Since η is pure and separable, it takes the form of a product vector

$$\eta \;=\; \left(\sum_{i=1}^{d} c_i\varphi_i\right) \otimes \left(\sum_{i=1}^{d} \tilde{c}_j\varphi_j\right) \;=\; \sum_{i,j=1}^{d} c_i\tilde{c}_j\varphi_i \otimes \varphi_j. \tag{2.4}$$

Comparing coefficients in (2.3) and (2.4), shows, that

$$c_1 \cdot \tilde{c}_1 \;= \alpha, \quad c_2 \cdot \tilde{c}_2 \;= \beta, \quad c_i \cdot \tilde{c}_j \;=\; 0 \text{ for } i \neq j \text{ or } i, j > 2. \tag{2.5}$$

As a consequence of the above equalities, $\alpha \cdot \beta = c_1 \cdot \tilde{c}_2 \cdot \tilde{c}_1 \cdot c_2 = 0$, which is a contradiction to $\alpha, \beta \neq 0$. We record

Example 13. *A pure bipartite state $\eta = \alpha \cdot \varphi_1 \otimes \varphi_1 + \beta \cdot \varphi_2 \otimes \varphi_2$, $|\alpha|^2 + |\beta|^2 = 1$ is entangled if and only if $\alpha, \beta \neq 0$.*

Schmidt decomposition

Especially useful is the following "polar" representation of vectors on a tensor product.

Theorem 14. *Let $a \in \mathcal{H}_1 \otimes \mathcal{H}_2$. Then there exist ortonormal systems $\{v_i\}_{i=1}^m \subset \mathcal{H}_1$, $\{w_i\}_{i=1}^m \subset \mathcal{H}_2$ and numbers $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m > 0$, such that*

$$a = \sum_{i=1}^{m} \sqrt{\alpha_i} v_i \otimes w_i$$

holds.

Remark 15. 1. *m is usually called the* Schmidt number *of a, $\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_m}$ the* Schmidt coefficients .

2. *It holds*

$$A_1 := \mathrm{tr}_{\mathcal{H}_2}(|a\rangle\langle a|) = \sum_{i=1}^{m} \alpha_i |v_i\rangle\langle v_i|$$

$$A_2 := \mathrm{tr}_{\mathcal{H}_1}(|a\rangle\langle a|) = \sum_{i=1}^{m} \alpha_i |w_i\rangle\langle w_i|.$$

In particular the Schmidt coefficients are the nonzero eigenvalues (counted with their multiplicities), and the vectors in the orthonormal systems which appear in the Schmidt decomposition are the corresponding eigenvectors.

Though formulated for general bipartite vectors, we will use the Schmidt decomposition most of the time for pure quantum states. The Schmidt decomposition Theorem is essentially a reformulation of the singular value decomposition, see Theorem 136.

Proof of Theorem 14. Using the inverse of the linear isomorphism Γ from (2.2), we obtain a matrix $A = \Gamma^{-1}(|a\rangle)$. Let

$$A = \sum_{i,j=1}^{m} \sqrt{\alpha_i}\, |v_i\rangle \overline{\langle w_i|}$$

be a singular value decomposition of A. Then

$$a \ = \Gamma \circ \Gamma^{-1} a = \sum_{i=1}^{m} \sqrt{\alpha_i}\, \Gamma(|w_i\rangle \overline{\langle v_i|}) = \sum_{i=1}^{m} \sqrt{\alpha_i} v_i \otimes w_i.$$

□

As we will see in forthcoming lectures, Schmidt decomposition is a very convenient way to write a pure bipartite state in proofs and calculations. Unfortunately, there seems to be no general satisfactory extension of Theorem 14 to spaces with more than two tensor factors. Nevertheless, Schmidt decompositions allows us to construct "purifications" of states, we define.

Definition 16 (Purification)**.** *Let $\rho \in \mathcal{S}(\mathcal{H})$ be a state, and $\mathcal{K}$ be an additional Hilbert space. The pure state $|\Psi\rangle\langle\Psi|$ is a* purification *of ρ if,* $\mathrm{tr}_{\mathcal{K}} |\Psi\rangle\langle\Psi| = \rho$.

The Schmidt decomposition offers a nice principle to construct purifications of a state ρ. By the remark following Theorem 14, the Schmidt coefficients and the orthonormal vectors in one tensor factor in Schmidt decompositions are obtained from a spectral decomposition of ρ. The remaining Schmidt vectors can be chosen freely on any Hilbert space of sufficient dimension (see Exercise 26).

Observation 17. *A quantum system in a pure state can only be uncorrelated to the "outside world", i.e. if $\rho \in \mathcal{S}(\mathcal{H} \otimes \mathcal{K})$ is such that $\mathrm{tr}_{\mathcal{K}}\rho$ is pure, it is necessarily a product state.*

The above observation also holds in for the classical statistical theory. However, in quantum information theory, this insight becomes a powerful tool when combined with the possiblity to "purify" quantum systems. The bipartite pure state resulting from purification captures all correlations of the system "to the outside world". This is a "quantum feature", since purifying systems is not possible in classical theory (see Exercise 28.)

2.1 Supplement: Entanglement witnesses

Definition 18 (Entanglement witness)**.** *A matrix $A \in \mathcal{L}^h(\mathcal{H}_A \otimes \mathcal{H}_B)$ is called an* entanglement witness *, if it is not positive semidefinite, and*

$$\mathrm{tr} A(\rho \otimes \sigma) \geq 0 \tag{2.6}$$

for all $\rho \in \mathcal{S}(\mathcal{H}_A)$, $\sigma \in \mathcal{S}(\mathcal{H}_B)$. An entangled state τ is said to be detected *by A, if*

$$\mathrm{tr} A\tau < 0. \tag{2.7}$$

Definition 19. *A hyperplane in $\mathbb{R}^m$ is a set of the form*

$$H(\xi, k) := \{z \in \mathbb{R}^m : \langle \xi, z\rangle = k\}, \tag{2.8}$$

where $\xi \in \mathbb{R}^m \setminus \{0\}$, and $k \in \mathbb{R}$. $H(\xi, k)$ is said to separate two sets $S_1, S_2 \subset \mathbb{R}^m$, if

Some properties

1. $H(\xi, k) \perp span\{\xi\}$
2. $\forall x, y \in H(\xi, k), \lambda \in \mathbb{R}$, it holds $\langle \lambda, \xi, x - y\rangle$

Theorem 20 (Separating Hyperplanes)**.** *If S is a closed convex subset of $\mathbb{R}^m$, and $x_0 \notin S$, then there exists a hyperplane separating x_0 and S.*

Proof. First step. We show, that there is a unique $s_0 \in S$ minimizing the distance $|s - x_0|$. Such an s_0 exists, because S is closed. Namely, choosing a sequence $\{s_i\}_{i\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} |s_k - x_0| = \inf_{s\in S} |s - x_0| \tag{2.9}$$

holds. By the Bolzano-Weierstrass Theorem ("Every bounded sequence has a convergent subsequence"), we find a subsequence $\{s_{k(j)}\}_{j=1}^{\infty}$, which converges to an $s_0 \in \mathbb{R}^m$, which is also a member of S, since by hypothesis, S is closed. Consequently,

$$|s_0 - x_0| \;=\; \lim_{j\to\infty} |s_{k(j)} - x_0| \;=\; \lim_{k\to\infty} |s_k - x_0| \;=\; \inf_{s\in S} |s - x_0|. \tag{2.10}$$

We show uniqueness by contradiction. Assume that another element $s' \in S$ also fulfills the mentioned condition, i.e.

$$|s' - x_0| \;=\; \inf_{s\in S} |s' - x_0|. \tag{2.11}$$

Then $|s' - x_0| = |s_0 - x_0|$, and x_0, s_0, s' form an isosceles triangle. But then the midpoint of $\overline{s_0 s'}$) (which is by convexity of S also a member of S), is closer to x_0 than s_0, a contradiction!
As a second step, we construct a hyperplane which separates s_0 and x_0. We set

$$\xi := x_0 - s_0, \text{ and } \quad \hat{\xi} := s_0 + \tfrac{1}{2}\xi. \tag{2.12}$$

We have

$$k \;:=\; \langle \xi, \hat{\xi}\rangle \;=\; \langle x_0 - s_0, \tfrac{1}{2}(x_0 + s_0)\rangle \;=\; \tfrac{1}{2}(|x_0|^2 - |s_0|^2). \tag{2.13}$$

By construction, $H(\xi, k)$, the hyperplane perpendicular to ξ, going through $\hat{\xi}$, fulfills

$$0 \;<\; \tfrac{1}{2}|s_0 - x_0|^2 = \frac{1}{2}(\langle \xi, s_0\rangle - \langle \xi, x_0\rangle) \tag{2.14}$$

Therefore, $H(\xi,k)$ is a separating Hyperplane for s_0 and x_0, since

$$\langle \xi, x_0\rangle \; < \frac{1}{2}(\langle \xi, s_0\rangle - \langle \xi, x_0\rangle) \; < \; \langle \xi, s_0\rangle. \tag{2.15}$$

In the last two steps of the proof, we convince ourselves, that $H(\xi,k)$ is indeed separating x_0 and S.
Step three. We show $H(\xi,k)\cap S = \emptyset$. Assume that $s_1 \in H(\xi,k)\cap S$, i.e. $\langle \xi, s_1\rangle = k$. Consider the isosceles triangle with vertices s_1, x_0, s_0. Let s_2 be a point on $\overline{s_0 s_1}$ for which $\overline{x s_2} \perp \overline{s_0 s_1}$. Then

$$|x_0 - s_2| \; < \; |x_0 - s_0|, \tag{2.16}$$

which is a contradiction, because $s_2 \in S$. consequently the constructed hyperplane and S do not intersect.
Fourth step. We show, again by contradiction, that $\langle \xi, s\rangle > k$ for all $s \in S$, i.e. $H(\xi,k)$ separates S from x_0. Assume hat $s_1 \in S$ and $\alpha_1 := \langle \xi, s_1\rangle \leq k$. We already know from Step 2, that $\alpha_0 := \langle \xi, s_0\rangle > k$. Define

$$\lambda \; := \; \frac{k-\alpha_1}{\alpha_0 - \alpha_1}. \tag{2.17}$$

Notice, that λ is in $[0,1)$, since

$$\alpha_0 \; > \; k \; \geq \alpha_1. \tag{2.18}$$

By convexity of S,

$$s_2 := \lambda s_0 + (1-\lambda)s_1 \; \in \; S, \tag{2.19}$$

which is a contradiction, because

$$\langle \xi, s_2\rangle = \lambda\alpha_0 + (1-\lambda)\alpha_1 = k. \tag{2.20}$$

□

Although we formulated the separating hyperplane Theorem for $\mathbb{R}^m$ and euclidean scalar product $\langle\cdot,\cdot\rangle$, there is no obstacle in using it for the set $\mathcal{L}(\mathcal{H})$ equipped with the Hilbert scalar product $\langle\cdot,\cdot\rangle_{HS}$ on that space. Indeed $\Gamma : \mathbb{R}^{2m^2} \to \mathcal{L}(\mathcal{H})$ with

$$A \; \mapsto (\mathfrak{Re}(a_{11}),\ldots,\mathfrak{Re}(a_{mm}),\mathfrak{Im}(a_{11}),\ldots,\mathfrak{Im}(a_{mm})) \tag{2.21}$$

sets up a linear isomorphism between $\mathbb{R}^{2m^2}$ and $\mathcal{L}(\mathcal{H})$, and it is easily checked, that

$$\langle A, B\rangle_{HS} \; = \; \langle \Gamma(A), \Gamma(B)\rangle \tag{2.22}$$

for all $A, B \in \mathcal{L}(\mathcal{H})$.

Theorem 21. *A state $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is*

- *separable, if and only if* $\mathrm{tr}\rho A \geq 0$ *for each entanglement witness A.*
- *entangled, if and only if there exists an entanglement witness A, such that* $\mathrm{tr}\rho A < 0$.

Proof. The both claims of the above theorem are easily seen to be equivalent. We prove the first claim. Asssume, that ρ is separable. By defnition, $\mathrm{tr}\rho A \geq 0$ for each entanglement witness A. for the converse statement, we notice, that the set of separable states on $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is a closed convex subset of $\mathcal{L}(\mathcal{H})$. By Theorem 20 and the remark after its proof, we find a Hyperplane $H(A,k) \subset \mathcal{L}(\mathcal{H})$, such that $\langle A, \sigma \rangle_{HS} \geq k$ for each member σ of the separable density matrix, and $\langle A, \rho \rangle < k$. Consequently, $A - k\mathbb{1}_{\mathcal{H}_A \otimes \mathcal{H}_B}$ is an entanglement witness for ρ. □

2.2 Exercises

Exercise 22 (Non-cyclicity of the partial trace)**.** *Find an example of matrices $A, B \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ such that $\mathrm{tr}_{\mathcal{H}_2}(AB) \neq \mathrm{tr}_{\mathcal{H}_2}(BA)$ holds.*

Exercise 23 (Compatibility problems)**.** *Let $q \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$, $r \in \mathcal{P}(\mathcal{X}_2 \times \mathcal{X}_3)$ be probability distributions. They are called* compatible*, if there is a probability distribution $p \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ having q and r as marginals on $\mathcal{X}_1 \times \mathcal{X}_2$ and $\mathcal{X}_2 \times \mathcal{X}_3$ respectively.*

- *Show, that q, r are compatible if and only if their marginals on the "shared alphabet" $\mathcal{X}_2$ coincide.*
- *Show, by counterexample, that an analogue of the above equivalence does not hold for density matrices. (Hint: Monogamy of entanglement).*

Exercise 24. *Mixing may cause destruction of entanglement. Convince yourself by straightforward calculation, that the equidistributed mixture*

$$\rho := \frac{1}{2}(|\Phi_+\rangle\langle\Phi_+| + |\Phi_-\rangle\langle\Phi_-|) \tag{2.23}$$

of the so-called singlet states with state vectors Φ_+ and Φ_-,

$$\Phi_\pm := \frac{1}{\sqrt{2}}(e_0 \otimes e_0 \pm e_1 \otimes e_1) \tag{2.24}$$

is in fact separable.

Exercise 25. *The set of separable density matrices in $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is convex by definition. What are the extremal elements?*

Exercise 26. *Let $\rho \in \mathcal{S}(\mathcal{H})$ be a quantum state. What is the minimum dimension of a Hilbert space $\mathcal{K}$ such that we find a purification of ρ on $\mathcal{S}(\mathcal{H} \otimes \mathcal{K})$?*

Exercise 27. *Let $\rho \in \mathcal{S}(\mathcal{H})$ be a density matrix with spectral decomposition $\sum_{i=1}^r \lambda_i |f_i\rangle\langle f_i|$. Show, that with a pure maximally entangled state $\Phi := |\phi\rangle\langle\phi|$ with state vector*

$$\phi := \frac{1}{\sqrt{r}} \sum_{i=1}^r f_i \otimes f_i$$

on $\mathcal{H} \otimes \mathcal{H}$,

$$\psi := (\mathbb{1}_{\mathcal{H}} \otimes \rho^{\frac{1}{2}})\phi$$

is state vector of a purification of ρ.

Exercise 28 (Classical purifications)**.** *Let $p \in \mathcal{P}(\mathcal{X})$ be mixed. There exists no $\mathcal{Y}$ such that $r \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is pure and the $\mathcal{X}$-marginal is p.*

3 Quantum hypotheses testing - Quantum Stein's Lemma

In this lecture we consider the "asymmetric quantum hypothesis testing" problem. Assume, an experimenter is confronted with a source which emits pairwise independent and equally prepared quantum systems. Given two a priori density matrices σ_0 (called "null hypothesis"), or σ_1 (called "alternative hypothesis") the goal is to decide by measurements on the outputs, which preparation is present. We prove "Quantum Stein's lemma", which quantifies the behaviour of the error of optimal tests for this task in a situation, where large numbers of outputs of the systems are available for performing tests.
This lecture is central to the course for two reasons. A first one is, that hypothesis tests also make up for good message transmission codes. This fact is already known from classical Shannon information theory. However, this relation seems to be even more important for quantum systems, as we will see in subsequent sections. A second reason is, that Stein's lemma allows a very simple and illuminating proof of the *monotonicity of the quantum relative entropy under completely positive and trace preserving maps*, which is notoriously hard to prove otherwise. After all, this will be our entrance to several highly notrivial quantum entropic inequalities which are essential for proving major results in quantum Shannon theory.
The mentioned strategy to prove entropy inequalities starting from Quantum Stein's lemma is strongly inspired by the paper [1], where also the relatively elementary proof of the result given below can be found. The interested reader also should consult that work. To formally settle the above described situation, assume, we are confronted with a preparation device which emits quantum systems pairwise uncorellated and additionally all being prepared according to the same density matrix. If this density matrix is γ, the mentioned properties of the preparation device ensure us, that the statistical behaviour of the joint quantum state of n systems ("*blocklength n*") prepared is described by the density matrix

$$\gamma^{\otimes n} := \underbrace{\gamma \otimes \cdots \otimes \gamma}_{n \text{ times}}$$

To settle the (asymetric) Hypothesis test problem, assume now, that the state γ is unknown to the receiver of the systems. The receiver is provided with two a priori hypotheses in form of density matrices ρ ("null hypothesis") and σ ("alternative hypothesis"), and tries by measurement on the outputs, to decide, which of these hypothesis to accept. For a given test $\{E_0, E_1\}$ on the n-fold output system, two kinds of errors can happen

1. **First kind error:** The actual density matrix is ρ, but σ is detected. This happens with probability

$$\mathrm{tr} E_1 \rho^{\otimes n} = \mathrm{tr}(\mathbb{1} - E_0)\rho^{\otimes n}$$

2. **Second kind error:** The density matrix is σ, but ρ is detected. This happens with probability

$$\mathrm{tr} E_0 \sigma^{\otimes n}$$

3.1 Quantum Stein's Lemma

A common goal now is, to determine the optimal asymptotical behaviour of the second kind error for tests whose first kind error is below a threshold $\epsilon \in (0,1)$. For this reason, we define for each $\epsilon \in [0,1]$

$$\beta_{\epsilon,n}(\rho,\sigma) := \inf\left\{\mathrm{tr}(a\sigma^{\otimes n}) : \ 0 \le a \le \mathbb{1}_{\mathcal{H}}^{\otimes n} \text{ and } \mathrm{tr}(a\rho^{\otimes n}) \ge 1-\epsilon\right\}.$$

To formulate the quantum version of Stein's Lemma, we need the following definition

Definition 29 (Quantum relative entropy)**.** *The* quantum relative entropy *of a pair* $(\rho,\sigma) \in \mathcal{S}(\mathcal{K}) \times \mathcal{S}(\mathcal{K})$ *is defined*

$$D(\rho\|\sigma) := \begin{cases} \mathrm{tr}(\rho(\log\rho - \log\sigma)) & \text{if } \ker\sigma \subset \ker\rho \\ +\infty & \text{otherwise} \end{cases}$$

The following Theorem is the quantum theoretic generalization to Stein's Lemma.

Theorem 30 (Quantum Stein's Lemma)**.** *Let* $\rho,\sigma \in \mathcal{S}(\mathcal{H})$ *density matrices with* $\ker\sigma \subset \ker\rho$*. For each* $\epsilon \in (0,1)$*, it holds*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_{\epsilon,n}(\rho,\sigma) \ = \ -D(\rho\|\sigma)$$

We will prove the above theorem in two portions. First, we prove the claim

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_{\epsilon,n}(\rho,\sigma) \ \le \ -D(\rho\|\sigma) \tag{3.1}$$

which implies, together with

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_{\epsilon,n}(\rho,\sigma) \ \ge \ -D(\rho\|\sigma) \tag{3.2}$$

the assertion of the theorem.

Types

In this paragraph, we introduce a certain instance of "typical projections". To proceed, we need the following definition

Definition 31 (von Neumann entropy)**.** *The* von Neumann entropy *of a density matrix* $\rho \in \mathcal{S}(\mathcal{H})$ *is defined by*

$$S(\rho) := -\mathrm{tr}(\rho\log\rho).$$

Remark 32. *Note, that* $\log A$ *is not defined, if is not of full rank. The above definition is to be understood using the convention* $0 \log 0 = 0$.

First, we introduce some notions, we will use. Let $\tau_1,\ \tau_2$ be density matrices on $\mathcal{H}$, such that $\ker \tau_2 \subset \ker \tau_1$. We set

$$M(\tau_1 \| \tau_2) := -\mathrm{tr}\tau_1 \log \tau_2,$$

which allows us to write

$$D(\tau_1 \| \tau_2) = \mathrm{tr}(\tau_1 \log \tau_1) - \mathrm{tr}(\tau_1 \log \tau_2) = -S(\tau_1) + M(\tau_1 \| \tau_2),$$

and

$$M(\tau_1 \| \tau_1) = S(\tau_1)$$

in case that $\tau_1 = \tau_2$ holds. Set $d := \dim \mathcal{H}$ and

$$\tau_2 = \sum_{x=1}^{d} \mu(x) |\phi_x\rangle\langle\phi_x|$$

a spectral decomposition of τ_2. For each $n \in \mathbb{N}$, we obtain

$$\begin{aligned}
\tau_2^{\otimes n} &= \left(\sum_{x=1}^{d} \mu(x) |\phi_x\rangle\langle\phi_x| \right)^{\otimes n} \\
&= \bigotimes_{i=1}^{n} \left(\sum_{x_1=1}^{d} \mu(x_i) |\phi_{x_i}\rangle\langle\phi_{x_i}| \right) \\
&= \sum_{x_1=1}^{d} \cdots \sum_{x_n=1}^{d} \mu(x_1) \cdots \mu(x_n) |\phi_{x_1}\rangle\langle\phi_{x_1}| \otimes \cdots \otimes |\phi_{x_n}\rangle\langle\phi_{x_n}| \\
&= \sum_{x^n \in [d]^n} \mu^n(x^n) |\phi_{x^n}\rangle\langle\phi_{x^n}|. \qquad (3.3)
\end{aligned}$$

The last equality is from introducing the notation

$$\mu^n(x^n) := \mu(x_1) \cdots \mu(x_n) = \prod_{i=1}^{n} \mu(x_i), \quad \text{and} \quad \phi_{x^n} := \phi_{x_1} \otimes \cdots \otimes \phi_{x_n}$$

for each $x^n = (x_1, \ldots, x_n) \in [d]^n$. In fact, $\{\phi_{x^n} :\ x^n \in [d]^n\}$ is an orthonormal basis in $\mathcal{H}^{\otimes n}$. The right hand side of Eq. (3.3) is a spectral decomposition of $\tau_2^{\otimes n}$. We define for each $\delta > 0$, $n \in \mathbb{N}$ the set

$$T_{\delta,n}(\tau_1, \tau_2) := \left\{ x^n \in [d]^n :\ 2^{-n(M(\tau_1\|\tau_2)+\delta)} < \mu^n(x^n) < 2^{-(M(\tau_1\|\tau_2)-\delta)} \right\} \qquad (3.4)$$

According to the set $T_{\delta,n}(\tau_1, \tau_2)$, we define the projector

$$p_{\delta,n}(\tau_1, \tau_2) := \sum_{x^n \in T_{\delta,n}(\tau_1,\tau_2)} |\phi_{x^n}\rangle\langle\phi_{x^n}|$$

The next lemma collects some useful properties of the above type of projector.

Lemma 33. *Let $\tau_1, \tau_2 \in \mathcal{S}(\mathcal{H})$, $\ker \tau_2 \subset \ker \tau_1$. For each $\delta > 0$, it holds with the abbreviation $p_n := p_{\delta,n}(\tau_1, \tau_2)$*

1. $p_n \tau_2^{\otimes n} = \tau_2^{\otimes n}$ *for all* $n \in \mathbb{N}$,
2. $p_n \tau_2^{\otimes n} p_n \leq 2^{-n(M(\tau_1\|\tau_2)-\delta)} p_n$,
3. $p_n \tau_2^{\otimes n} p_n \geq 2^{-n(M(\tau_1\|\tau_2)+\delta)} p_n$,
4. $\lim_{n\to\infty} \mathrm{tr} p_n \tau_1^{\otimes n} = 1$.

Proof. The first three claims are obvious from the definitions. We will now show, that the fourth claim follows from an application of the law of large numbers. Define an i.i.d. sequence $X_1, \ldots, X_n$ of random variables with values in $[d]$ and probabilities

$$\Pr(X_i = x) = \langle \phi_x, \tau_1 \phi_x \rangle$$

for each $i \in [n]$. We consequently have

$$\begin{aligned} \Pr(X_1 = x_1 \wedge \cdots \wedge X_n = x_n) &= \prod_{i=1}^{n} \Pr(X_i = x_i) \\ &= \prod_{i=1}^{n} \langle \phi_{x_i}, \tau_1 \phi_{x_i} \rangle \\ &= \langle \phi_{x^n}, \tau_1^{\otimes n} \phi_{x^n} \rangle \end{aligned}$$

for each $x^n \in [d]^n$. Using the function $f : [d] \to \mathbb{R}$ defined by

$$f(x) := -\log \mu(x) \qquad (x \in [d]),$$

we obtain another i.i.d. sequence $U_1, \ldots, U_n$ of random variables, each defined by

$$U_i := f \circ X_i. \qquad (i \in [n]).$$

It holds

$$\begin{aligned} \mathbb{E} U_i &= \sum_{x\in[d]} \Pr(X_i = x) f(x) \\ &= -\sum_{x\in[d]} \Pr(X_i = x) \log \mu(x) \\ &= -\sum_{x\in[d]} \Pr(X_i = x) \langle \phi_x, \log \tau_2 \phi_x \rangle \\ &= -\sum_{x\in[d]} \langle \phi_x, \tau_1 \phi_x \rangle \langle \phi_x, \log \tau_2 \phi_x \rangle \\ &= -\mathrm{tr}(\tau_1 \log \tau_2) \\ &= M(\tau_1 \| \tau_2). \end{aligned}$$

Thus, it holds

$$\begin{aligned}\mathrm{tr}(\tau_1^{\otimes n} p_n) &= \sum_{x^n \in T_{\delta,n}(\tau_1,\tau_2)} \mathrm{tr}(\tau_1^{\otimes n} |\phi_{x^n}\rangle\langle\phi_{x^n}| \\ &= \Pr\big((X_1,\ldots,X_n) \in T_{\delta,n}(\tau_1,\tau_2)\big) \\ &= \Pr\left(\left|\frac{1}{n}\sum_{i=1}^n U_i - M(\tau_1\|\tau_2)\right| < \delta\right) \\ &= \Pr\left(\left|\frac{1}{n}\sum_{i=1}^n U_i - n\mathbb{E}(U_1)\right| < \delta\right)\end{aligned}$$

Taking the limit $n \to \infty$ in the chain of eqalities above, we obtain the fourth claim of the lemma by using the law of large numbers. □

Proof - Achievability

To prove achievability in Quantum Stein's lemma, i.e. the inequality in Eq. (3.1), we will define suitable tests which are essentially products of the projections $p_{n,\delta}(\rho,\sigma)$ and $p_{n,\delta}(\rho,\rho)$ for some fixed δ. These projections do not necessarily commute, therefore, we will need the following lemma.

Lemma 34. *Let $p, q \in \mathcal{L}(\mathcal{K})$, $0 \le p, q \le \mathbb{1}_{\mathcal{K}}$. It holds*

$$\mathrm{tr}(\tau pqp) \;\ge\; \mathrm{tr}(\tau q) - 2\sqrt{\mathrm{tr}\tau(\mathbb{1}_{\mathcal{K}} - p)}.$$

Proof. We abbreviate $\mathbb{1}_{\mathcal{K}}$ with $\mathbb{1}$. It holds

$$0 \;\le\; (\mathbb{1}-p)q(\mathbb{1}-p) \;=\; -q + q(\mathbb{1}-p) + (\mathbb{1}-p)q + pqp,$$

which is equivalent to

$$q \;\le\; q(\mathbb{1}-p) + (\mathbb{1}-p)q + pqp$$

which in turn, by multiplying with τ and using the conjugation rule Lemma 139 can be transformed to

$$\tau^{\frac{1}{2}} q \tau^{\frac{1}{2}} \;\le\; \tau^{\frac{1}{2}} q(\mathbb{1}-p)\tau^{\frac{1}{2}} + \tau^{\frac{1}{2}}(\mathbb{1}-p)q\tau^{\frac{1}{2}} + \tau^{\frac{1}{2}} pqp \tau^{\frac{1}{2}}.$$

Note that

$$\begin{aligned}\mathrm{tr}\tau^{\frac{1}{2}} q(\mathbb{1}-p)\tau^{\frac{1}{2}} + \mathrm{tr}\tau^{\frac{1}{2}}(\mathbb{1}-p)q\tau^{\frac{1}{2}} \;&=\; \mathrm{tr}\tau^{\frac{1}{2}} q(\mathbb{1}-p)\tau^{\frac{1}{2}} + \overline{\mathrm{tr}\tau^{\frac{1}{2}}(\mathbb{1}-p)q\tau^{\frac{1}{2}}} \\ &= 2\cdot Re\,\mathrm{tr}\tau^{\frac{1}{2}} q(\mathbb{1}-p)\tau^{\frac{1}{2}} \\ &\le 2\cdot|\mathrm{tr}\tau^{\frac{1}{2}} q(\mathbb{1}-p)\tau^{\frac{1}{2}}| \\ &\le 2\sqrt{\mathrm{tr}\tau q^2}\cdot\sqrt{\mathrm{tr}\tau(1-p)^2}.\end{aligned}$$

Using monotonicity of the trace, we arrive at

$$\begin{aligned}\mathrm{tr}\tau q \leq\ & \mathrm{tr}\tau pqp + \mathrm{tr}\tau^{\frac{1}{2}} q(\mathbb{1}-p)\tau^{\frac{1}{2}} + \mathrm{tr}\tau^{\frac{1}{2}}(\mathbb{1}-p)q\tau^{\frac{1}{2}} \\ \leq\ & \mathrm{tr}\tau pqp + 2\sqrt{\mathrm{tr}\tau q^2}\cdot\sqrt{\mathrm{tr}\tau(1-p)^2}.\end{aligned}$$

Consequently, we have

$$\mathrm{tr}\tau pqp \ \geq\ \mathrm{tr}\tau q - 2\sqrt{\mathrm{tr}\tau(\mathbb{1}-p)},$$

as we desired to prove. □

Proof of achievability in Theorem 30, i.e. Eq. (3.1). We show, for $\rho,\sigma\in\mathcal{S}(\mathcal{H})$, $\ker\sigma\subset\ker\rho$, $\epsilon>0$

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta_{\epsilon,n}(\rho,\sigma)\ \leq\ -D(\rho\|\sigma)$$

holds Fix numbers $\delta>0$, $n\in\mathbb{N}$. We define according to Lemma 33 $p_n := p_{\delta,n}(\rho,\rho)$, $q_n := p_{\delta,n}(\rho,\sigma)$ and an effect $a_n := q_n p_n q_n(\rho,\sigma)$, which will serve as the hypothesis test for blocklength n. It holds

$$\mathrm{tr}a_n\rho^{\otimes n}\ =\ \mathrm{tr}q_n p_n q_n\rho^{\otimes n}\ \geq\ \mathrm{tr}p_n\rho^{\otimes n} - 2\cdot\left(\mathrm{tr}(\mathbb{1}_{\mathcal{H}}^{\otimes n}-q_n)\rho^{\otimes n}\right)^{\frac{1}{2}}, \tag{3.5}$$

where Lemma 34 was used. Each of the summands in Eq. (3.5) can be upper bounded by using properties of the projections stated in Lemma 33 with $\tau_1=\tau_2=\rho$. We have by Lemma 33.4 applied with $\tau_1=\tau_2=\rho$

$$\lim_{n\to\infty}\mathrm{tr}p_n\rho^{\otimes n}=1, \tag{3.6}$$

and with $\tau_1=\rho$, $\tau_2=\sigma$

$$\lim_{n\to\infty}\mathrm{tr}p_n\rho^{\otimes n}=1. \tag{3.7}$$

Combination of (3.6) and (3.7) with (3.5) leads us to $\lim_{n\to\infty}\mathrm{tr}a_n\rho^{\otimes n}=1$. For each large enough n, we therefore have

$$\mathrm{tr}(a_n\rho^{\otimes n})\geq 1-\epsilon, \tag{3.8}$$

i.e. a_n belongs to the feasible set of optimization for $\beta_{\epsilon,n}(\rho,\sigma)$. We will show, that the second kind error of a_n is bounded in a suitable way for large enough blocklengths. Note, that

$$\mathrm{tr}p_n q_n p_n\ \leq\ \mathrm{tr}p_n\ \leq\ 2^{-n(S(\rho)-\delta)}\cdot\mathrm{tr}p_n\rho^{\otimes n}. \tag{3.9}$$

Consequently, it holds

$$\begin{aligned}\mathrm{tr}a_n\sigma^{\otimes n}\ =\ & \mathrm{tr}q_n\sigma^{\otimes n}q_n p_n \\ \leq\ & 2^{-n(M(\rho\|\sigma)-\delta)}\cdot\mathrm{tr}p_n q_n p_n \\ \leq\ & 2^{-n(D(\rho\|\sigma)-2\delta)}.\end{aligned}$$

Finally, we conclude, that for large enough n,

$$\beta_{n,\epsilon}(\rho,\sigma) \;\leq\; tra_n\sigma^{\otimes n} \leq \exp(-n(D(\rho\|\sigma)-2\delta)),$$

i.e.

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta_{\epsilon,n}(\rho,\sigma) \;\leq\; -D(\rho\|\sigma)+2\delta.$$

Since δ was an arbitrary positive number, we are done. □

Proof - Converse

For proving the converse statement (3.2), we will employ the following inequality, which can be derived from the one given in Lemma 34.

Lemma 35. *Let $p,q,u\in\mathcal{L}(\mathcal{K})$, $0\leq p,q\leq\mathbb{1}_{\mathcal{K}}$, $\tau\in\mathcal{S}(\mathcal{K})$, and u a projection with $u\tau=\tau u$, and $\tau u\leq cu$ for some $c\in\mathbb{R}_+$. Then*

$$\mathrm{tr}(pqp) \;\geq\; \frac{1}{c}\Big(\mathrm{tr}\tau q-2(\mathrm{tr}\tau(\mathbb{1}_{\mathcal{K}}-p))^{\frac{1}{2}}-\mathrm{tr}\tau(\mathbb{1}_{\mathcal{K}}-u)\Big).$$

Proof. It holds

$$\begin{aligned}\mathrm{tr}pqp &= \mathrm{tr}upqp+tr(\mathbb{1}-u)pqp\\ &\geq \mathrm{tr}upqp\\ &\geq \frac{1}{c}\,\mathrm{tr}\tau upqp\\ &= \frac{1}{c}\left(\mathrm{tr}\tau pqp-\mathrm{tr}\tau(\mathbb{1}-u)pqp\right)\\ &\geq \frac{1}{c}\left(\mathrm{tr}\tau pqp-\mathrm{tr}\tau(\mathbb{1}-u)pqp\right).\end{aligned}$$

Applying Lemma 34 to lower-bound $\mathrm{tr}\tau pqp$, we obtain the desired inequality. □

Proof of the converse to Quantum Stein's lemma. We fix $\epsilon\in(0,1)$ and show the inequality

$$\liminf_{n\to}\frac{1}{n}\log\beta_{\epsilon,n}(\rho,\sigma) \;\geq\; -D(\rho\|\sigma).$$

Fix $\gamma>1$. Let for $n\in\mathbb{N}$, $a_n\in[0,\mathbb{1}_{\mathcal{H}}^{\otimes n}]$ be any test, such that

$$\begin{aligned}\mathrm{tr}q_n\rho^{\otimes n} &\geq 1-\epsilon,\;\text{ and}\\ \mathrm{tr}q_n\sigma^{\otimes n} &\leq \gamma\cdot\beta_{\epsilon,n}\end{aligned}$$

are simultaneously fulfilled (note, that such a text always exists). It holds (with p_n,q_n as defined in the achievability proof)

$$\sigma^{\otimes n} \;\geq\; q_n\sigma^{\otimes n}q_n \geq \exp\left(-n(M(\rho\|\sigma)+\delta)\right)q_n,$$

by Lemma 3 applied with $\tau_1=\rho$, and $\tau_2=\sigma$. Consequently,

$$a_n^{\frac{1}{2}}\sigma^{\otimes n}a_n^{\frac{1}{2}} \;\geq\; \exp\left(-n(M(\rho\|\sigma)+\delta\right)a_n^{\frac{1}{2}}q_na_n^{\frac{1}{2}} \tag{3.10}$$

does hold, where the first inequality is by the conjugation rule (Lemma 139). Taking traces on both sides of the inequality in (3.10), we arrive at

$$\beta_{\epsilon,n}(\rho,\sigma) \;\geq\; \gamma^{-1}\mathrm{tr}a_n\sigma^{\otimes n} \;\geq\; \gamma^{-1}\cdot\exp\left(-n(M(\rho\|\sigma)+\delta)\cdot\mathrm{tr}a_nq_n\right.$$

Using Lemma 33 with $\tau_1=\tau_2=\rho$ leads us to

$$p_n\rho^{\otimes n}p_n \;\geq\; 2^{-n(S(\rho)-\delta)}p_n. \tag{3.11}$$

With the inequality in (3.11), the conditions of Lemma 35 are fulfilled with the assignments $u=p=q_n$, $q=a_n$. $\tau=\rho$, and $c=\exp(-n(S(\rho)-\delta))$. It holds

$$\begin{aligned}\mathrm{tr}q_na_nq_n \;\geq\; &\exp(n(S(\rho)-\delta))\\ &\cdot\left(\mathrm{tr}\rho^{\otimes n}a_n-2\left(\mathrm{tr}\rho^{\otimes n}(\mathbb{1}_{\mathcal{H}}^{\otimes n}-q_n)\right)^{\frac{1}{2}}-\mathrm{tr}\rho^{\otimes n}(\mathbb{1}_{\mathcal{H}}^{\otimes n}-q_n)\right).\end{aligned}$$

Note, that the first term in brackets on the r.h.s. of the above inequality approaches one, while the others go to zero. We conclude, that for large enough n, we have

$$\mathrm{tr}q_na_nq_n \;\geq\; \frac{1-\epsilon}{2}\cdot\exp(n(S(\rho)-\delta)).$$

Putting everything together and taking logarithms, we arrive at

$$\frac{1}{n}\log\beta_{\epsilon,n}(\rho,\sigma) \;\geq\frac{1}{n}\log\frac{1-\epsilon}{2\gamma}\cdot(-D(\rho\|\sigma)-2\delta)$$

Taking the limes inferior on both sides of the above inequality gives the desired result. □

3.2 Exercises

Exercise 36. *Use the strategy of the above proof to Stein's Lemma to provide yourself with a proof of the classical version of Stein's lemma. Hint: In the classical case, the Lemmas 34 and Lemma 35 can be replaced by usage of the union bound.*

4 Quantum Channels

When we introduced the basic steps which make up for a statistical experiment in Chapter 1, we left out a very important basic building block – evolution. Systems undergo state changes which alter their statistical properties before being registered. Such *re-preparations* can result from an environmental influence ("noise"), or be an effect of a intentional modification ("processing"). In this lecture, we identify the general class if maps which represent such changes of a systems state. First we remember the "evolutions" encountered in classical information theory.

Example 37. *In classical information theory, changes of the classical states ("classical channels") are usually described by* stochastic matrices *. A stochastic matrix $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ is an array $W = (W(y|x))_{x\in\mathcal{X},y\in\mathcal{Y}}$ such that $W(\cdot|x)$ is a probability distribution on $\mathcal{Y}$ for each $x \in \mathcal{X}$. The set of stochastic matrices is convex with the extremal elements being the permutation matrices.*

Heuristically speaking, we should demand from a "quantum channel", that it is linear and it should "map density matrices to density matrices". But things are not that simple, as we will see. In the next definition, we will use the map

$$\mathrm{id}_{\mathbb{C}^l} : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{K}), \quad \mathrm{id}_{\mathcal{K}}(A) = A \qquad (A \in \mathcal{L}(\mathcal{K}))$$

Definition 38 (c.p. map)**.** *A linear map $\mathcal{T} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ is called*

1. positive *if for each $A \in \mathcal{L}(\mathcal{H})$, $A \geq 0$ also $\mathcal{T}(A) \geq 0$ holds.*
2. completely positive (c.p.) *if for each $l \in \mathbb{N}$ the map $id_{\mathcal{L}(\mathbb{C}^l)} \otimes \mathcal{T}$ is positive.*

At first sight, the distinction between positive and completely positive maps made above seems obsolete. As demonstrated by the following example, there are indeed maps which are positive – but not completely!

Example 39 (Partial transposition)**.** *The map $\mathcal{E} : \mathcal{L}(\mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2)$, $\mathcal{E}(A) := A^T$ (the transposition being according to the canonical orthonormal basis) is positive, but not completely positive.*

We note, that from an operational point of view, we have to demand complete positivity rather than positivity when defining what a quantum channel is. Consider for instance a situation, where a system A is processed with a map $\mathcal{T}$ in a lab, while the overall state ρ_{AE} of the composite system AE including an additional environment system is not a product state (i.e. A is an "open system"). If $\mathcal{T}$ is allowed to be an positive, but not c.p. map, it happens, that the global resulting state $\rho'_{AE} := \mathrm{id}_E \otimes \mathcal{T}(\rho)$ may not be a density matrix!

Definition 40 (Quantum channel)**.** *A linear map $\mathcal{T} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ is a called a* quantum channel *, if it is completely positive trace preserving (i.e. for each $A \in \mathcal{L}(\mathcal{H})$, $\mathrm{tr}\mathcal{T}(A) = \mathrm{tr}A$). The set of quantum channels (or "c.p.t.p. maps") with input space $\mathcal{H}$ and output space $\mathcal{K}$ is denoted $\mathcal{C}(\mathcal{H}, \mathcal{K})$.*

A first set of prominent examples of quantum channels are

a.) **Isometric evolutions**. With an isometry $v \in \mathcal{L}(\mathcal{H})$, the map

$$\mathcal{V}(a) := vav^* \qquad (a \in \mathcal{L}(\mathcal{H})) \tag{4.1}$$

is a channel.

b.) **Partial traces**. The map $\mathrm{tr}_{\mathcal{K}} : \mathcal{L}(\mathcal{H} \otimes \mathcal{L}) \to \mathcal{L}(\mathcal{H})$ as defined in the previous lecture is a channel.

c.) **Appending**. The map $\mathcal{N}_b : \mathcal{L}(\mathcal{H}) \otimes \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$, $a \mapsto a \otimes b$ is a channel for each $b \in \mathcal{S}(\mathcal{K})$, $b \geq 0$.

Prominent examples for maps which are **not** channels are

d.) **Transposition** (is not c.p., see Example 39)

e.) **Universal quantum copying device**. The map $T : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$, $\rho \mapsto \rho \otimes \rho$ is not a channel since it in not linear.

Remark 41. *The unavailability of a universal quantum copying devices has severe impact on the conception of quantum communication systems. On one hand, it allows for very powerful protocols protecting information from eavesdropping. On the other hand, a "quantum internet" needs completely novel solutions to the problem of long-distance transmission. Read-and-repeat solutions as performed by nowadays "repeater stations" are "impossible machines" for quantum transmission!*

From first sight, the condition of complete positivity formulated in Definition 40 seems to be not very handy. To check if a map $\mathcal{T}$ is completely positive, one would have to check if $\mathrm{id}_{\mathbb{C}^l} \otimes \mathcal{T}$ is positive for all $l \in \mathbb{N}$ – a rather hopeless task. Fortunately, there exist characterizations which are more easy to handle. The first one is given in the next theorem.

Theorem 42 (Kraus decomposition)**.** *Let $n := \dim \mathcal{H}$, $m := \dim \mathcal{K}$. A linear map $\mathcal{T} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ is completely positive if and only if it admits a representation of the form*

$$\mathcal{T}(B) = \sum_{k=1}^{N} T_k B T_k^* \qquad (B \in \mathcal{L}(\mathcal{H})) \tag{4.2}$$

with $T_1, \ldots, T_N \in \mathcal{L}(\mathcal{H}, \mathcal{K})$. A representation as (4.2) is then always possible with $N \leq n^2 \cdot m^2$. Moreover, $\mathcal{T}$ is c.p.t.p. if and only if it is c.p. and $\sum_{k=1}^{N} T_k^* T_k = \mathbb{1}_n$ holds.

Proof. We use the abbreviation $\mathbb{M}_n := \mathbb{M}_{n \times n}(\mathbb{C})$ We conduct our arguments mainly with matrices, i.e. $\mathbb{M}_n = \mathcal{L}(\mathcal{H})$, $\mathbb{M}_m = \mathcal{L}(\mathcal{K})$. First we show that if $\mathcal{T}$ admits a representation as in (4.2), it is completely positive. Let $l \in \mathbb{N}$, $C \in \mathbb{M}_l \otimes \mathbb{M}_n$ a positive semidefinite matrix. From (4.2) we obtain the equality

$$\mathrm{id}_{\mathbb{C}^l} \otimes \mathcal{T}(C) = \sum_{k=1}^{N} \mathbb{1}_{\mathbb{C}^l} \otimes T_k C (\mathbb{1}_{\mathbb{C}^l} \otimes T_k)^*. \tag{4.3}$$

Notice, that $\mathbb{1}_{\mathbb{C}^l} \otimes T_k X (\mathbb{1}_{\mathbb{C}^l} \otimes T_k)^*$ is p.s.d. by positivity of C and an application of the conjugation rule. Consequently $\mathrm{id}_{\mathbb{C}^l} \otimes \mathcal{T}$ is a positive map by (4.3).
We show the backwards implication (the possibility to represent $\mathcal{T}$ as in (4.2) if it is c.p.) using the canonical orthonormal basis $\{|e_i\rangle\langle e_i|\}$, since if we show

$$\mathcal{T}(|e_i\rangle\langle e_j|) = \sum_{k=1}^{N} T_k |e_i\rangle\langle e_j| T_k^* \tag{4.4}$$

for all $i, j \in [m]$ this implies the desired inequality by linearity. We use the linear isomorphism between $\mathbb{M}_m \otimes \mathbb{M}_n$ and the space $\mathbb{M}_n(\mathbb{M}_m)$ of block matrices with $n \times n$ blocks of size $m \times m$. This isomorphism is given by the map

$$\Lambda : \mathbb{M}_n \otimes \mathbb{M}_m \to \mathbb{M}_n(\mathbb{M}_m),$$

$$|e_i\rangle\langle e_j| \otimes |e_l\rangle\langle e_r| \;\mapsto\; \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ & \cdots & \vdots & \cdots & 0 \\ & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & |e_i\rangle\langle e_j| & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}, \tag{4.5}$$

the nonzero block matrix entry above is in the $(l, r,)$ block coordinate. We consider the so-called *Choi matrix* $C(\mathcal{T})$ of $\mathcal{T}$,

$$C(\mathcal{T}) \;=\; (\mathrm{id}_{\mathbb{C}^n} \otimes \mathcal{T}) \left[\sum_{i,j=1}^{n} |e_i\rangle\langle e_j| \otimes |e_i\rangle\langle e_j| \right] \;=\; \sum_{i,j=1}^{n} |e_i\rangle\langle e_j| \otimes \mathcal{T}(|e_i\rangle\langle e_j|) \tag{4.6}$$

which has, written as a block matrix, the form

$$\Lambda(C(\mathcal{T})) \;=\; \begin{pmatrix} \mathcal{T}(|e_1\rangle\langle e_1|) & \cdots & \mathcal{T}(|e_1\rangle\langle e_n|) \\ \vdots & \ddots & \vdots \\ \mathcal{T}(|e_n\rangle\langle e_1|) & \cdots & \mathcal{T}(|e_n\rangle\langle e_n|) \end{pmatrix}. \tag{4.7}$$

The matrix

$$\sum_{i,j=1}^{n} |e_i\rangle\langle e_j| \otimes |e_i\rangle\langle e_j| \tag{4.8}$$

is up to a positive scalar prefactor an orthogonal projector, and therefore positive semidefinite. Since $\mathcal{T}$ is assumed to be completely positive, $C(\mathcal{T})$ is also p.s.d., and we can write $C(\mathcal{T})$ in spectral decomposition

$$C(\mathcal{T}) \;=\; \sum_{i=1}^{N} \lambda_i |f_i\rangle\langle f_i| = \sum_{i=1}^{N} |v_i\rangle\langle v_i|. \tag{4.9}$$

with $N = \text{rank}C(\mathcal{T})$, and $v_i := \sqrt{\lambda_i} f_i$. We write

$$v_i \;=\; \sum_{j=1}^{n}\sum_{k=1}^{m} x_{jk}^{(i)}\; e_i \otimes e_k$$

for each $i \in [N]$. Therefore, we have

$$\begin{aligned}
|v_i\rangle\langle v_i| \;&=\; \sum_{j,l=1}^{n}\sum_{k,r=1}^{m} x_{jk}^{(i)}\overline{x}_{lo}^{(i)}\,|e_j\rangle\langle e_l| \otimes |e_k\rangle\langle e_r| \\
&=\; \sum_{j,l=1}^{n} |e_j\rangle\langle e_l| \otimes \left(\sum_{k,r=1}^{m} x_{jk}^{(i)}\overline{x}_{lr}^{(i)}\,|e_k\rangle\langle e_r|\right) \\
&=\; \sum_{j,l=1}^{n} |e_j\rangle\langle e_l| \otimes |x_j^{(i)}\rangle\langle x_l^{(i)}|,
\end{aligned}$$

where we used the definition

$$x_j^{(i)} := \sum_{k=1}^{n} x_{jk}^{(i)} e_k.$$

We have,

$$\Lambda(|v_i\rangle\langle v_i|) \;=\; \begin{pmatrix} |x_1^{(i)}\rangle\langle x_1^{(i)}| & \cdots & |x_1^{(i)}\rangle\langle x_n^{(i)}| \\ \vdots & \ddots & \vdots \\ |x_n^{(i)}\rangle\langle x_1^{(i)}| & \cdots & |x_n^{(i)}\rangle\langle x_n^{(i)}| \end{pmatrix}.$$

If we now define matrices $T_i = \begin{pmatrix} x_1^{(i)} \cdots x_n^{(i)} \end{pmatrix}$, it holds $T_i e_r = x_r^{(i)}$, i.e.

$$T_i\,|e_r\rangle\langle e_s|\,T_i^* \;= |x_r^{(i)}\rangle\langle x_s^{(i)}|.$$

Therefore, we have

$$\begin{aligned}
\Lambda(C(\mathcal{T})) \;&=\; \begin{pmatrix} \mathcal{T}(|e_1\rangle\langle e_1|) & \cdots & \mathcal{T}(|e_1\rangle\langle e_n|) \\ \vdots & \ddots & \vdots \\ \mathcal{T}(|e_n\rangle\langle e_1|) & \cdots & \mathcal{T}(|e_n\rangle\langle e_n|) \end{pmatrix} \\
&= \sum_{i=1}^{N} \Lambda(|v_i\rangle\langle v_i|) \\
&= \sum_{i=1}^{N} \begin{pmatrix} |x_1^{(i)}\rangle\langle x_1^{(i)}| & \cdots & |x_1^{(i)}\rangle\langle x_n^{(i)}| \\ \vdots & \ddots & \vdots \\ |x_n^{(i)}\rangle\langle x_1^{(i)}| & \cdots & |x_n^{(i)}\rangle\langle x_n^{(i)}| \end{pmatrix} \\
&= \sum_{i=1}^{N} \begin{pmatrix} T_i\,|e_1\rangle\langle e_1|\,T_i^* & \cdots & T_i\,|e_1\rangle\langle e_n|\,T_i^* \\ \vdots & \ddots & \vdots \\ T_i\,|e_n\rangle\langle e_1|\,T_i^* & \cdots & T_i\,|e_n\rangle\langle e_n|\,T_i^* \end{pmatrix}
\end{aligned}$$

Comparing the coefficients of the matrices on both ends of the chain of equalities above shows (4.4).
It remains to prove the second claim of the theorem. If $\mathcal{T}$ is c.p.t.p. , it holds for each $i, j \in [n]$

$$\begin{aligned}\delta_{ij} &= \operatorname{tr}|e_i\rangle\langle e_j| \\ &= \operatorname{tr}\mathcal{T}(|e_i\rangle\langle e_j|) \\ &= \operatorname{tr}\left(\sum_{k=1}^{N} T_k|e_i\rangle\langle e_j|T_k^*\right) \\ &= \sum_{k=1}^{N}\operatorname{tr}(T_k^*T_k|e_i\rangle\langle e_j|) \\ &= \langle e_j, \sum_{k=1}^{N} T_k^*T_k e_i\rangle.\end{aligned}$$

□

Next, we give a second characterization of completely positive maps. While a Kraus decomposition is a representation "on a lower layer" (a map between matrix spaces is represented by a matrix sum, the theorem below gives a representation on a "higher layer". A c.p. map is represented by a partial evolution derived from an isometric evolution on a larger space.

Theorem 43 (Stinespring dilation)**.** *Let $\mathcal{T} \in \mathcal{C}(\mathcal{H}, \mathcal{H}')$ be a c.p. map. Then exists a representation of $\mathcal{T}$ in terms of a linear map $v : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}' \otimes \mathcal{K})$ with additional Hilbert space $\mathcal{K}$ such that*

$$\mathcal{T}(a) = \operatorname{tr}_{\mathcal{K}}(vav^*) \qquad (a \in \mathcal{L}(\mathcal{H})). \tag{4.10}$$

holds. If $\mathcal{T}$ is c.p.t.p., v is an isometry.

Proof. Since $\mathcal{T}$ is completely positive there exists a Kraus decomposition with Kraus operators $T_1, \dots T_N$, $T_i \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$. Set $\mathcal{H}_K := \mathbb{C}^N$ and define

$$v = \sum_{k=1}^{N} T_k \otimes |e_k\rangle. \tag{4.11}$$

We convince ourselves, that v is in fact an isometry when $\mathcal{T}$ is c.p.t.p. . We have to show, that

$$v^*v = \mathbb{1}_{\mathcal{H}} \quad \text{and} \quad vv^* \tag{4.12}$$

is a projection onto a subspace of $\mathcal{H}' \otimes \mathcal{K}$. This can be verified as follows. We have

$$v^*v = \sum_{k,l=1}^{N} T_k^*T_l \cdot \langle e_k, e_l\rangle = \sum_{k=1}^{N} T_k^*T_k = \mathbb{1}_{\mathcal{H}}. \tag{4.13}$$

On the other hand, vv^* is clearly hermitian. That it is also idempotent can be seen directly via

$$(vv^*)^2 = v \underbrace{v^*v}_{=\mathbb{1}_{\mathcal{H}}} v^* = vv^*. \tag{4.14}$$

□

4.1 Exercises

Exercise 44. *The* Hilbert Schmidt adjoint $\mathcal{N}_*$ *of a map* $\mathcal{N} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ *is the map, which fulfills*

$$\mathrm{tr}(\mathcal{N}(A^*)B) = \mathrm{tr}(A^*\mathcal{N}_*(B)) \tag{4.15}$$

for all $A \in \mathcal{L}(\mathcal{H})$, $B \in \mathcal{L}(\mathcal{K})$. *Show the following*

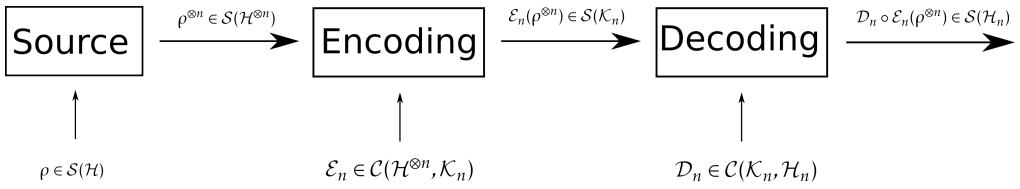
1. *Complete positivity of* $\mathcal{N}$ *implies complete positivity of* $\mathcal{N}_*$.
2. *If* $\mathcal{N}$ *is moreover trace preserving,* $\mathcal{N}_*$ *has an additional property, which one?*

5 Source compression for memoryless quantum sources

We consider a *discrete memoryless quantum source* (DMQS) with generic density matrix $\rho \in \mathcal{S}(\mathcal{H})$. In this model, the state of n source outputs is the n-fold tensorial extension of the generic density matrix ρ, i.e.

$$\rho^{\otimes n} = \underbrace{\rho \otimes \cdots \otimes \rho}_{n \text{ times}}.$$

In this lecture we consider the question to which extend the outputs of such a source can be compressed. We regard "degrees of freedom" i.e. "Hilbert space dimensions" as a costly resource and seek for source compression schemes, which map the statistics of the source outputs to a system with smaller number of dimensions such that it can be recovered in a sufficient way.
In case, that we aim to perfectly store n outputs of the DMQS ρ, one can show, that a space of $(\text{rank}\rho)^n$ dimensions is necessary. This amount can sometimes be decreased substantially, if we allow small imprecisions in recovery. A the general scheme of a source compression code for blocklength n is depicted below.



In order to show a coding theorem and a converse to quantify the asymptotics of the optimal compression rates, we need a sufficient performance measure to quantify the quality of recovery. We will use the *entanglement fidelity* for this purpose.
In Section 5.1 we introduce (quantum) fidelity and entanglement fidelity. In Section 5.2, we state and prove the source compression theorem for DMQS. Therein, the optimal compression rate will be determined by the von Neumann entropy of the generic density matrix. We discuss some properties of this function in Section 5.3.

5.1 Fidelity, and Entanglement Fidelity

Definition 45 (Quantum fidelity). *Let $A, B \in \mathcal{L}(\mathcal{K})$, $A, B \geq 0$. The (quantum) fidelity of A and B is defined by*

$$F(A,B) := \|A^{\frac{1}{2}} B^{\frac{1}{2}}\|_1^2 = \left(\text{tr}\left[(A^{\frac{1}{2}} B A^{\frac{1}{2}})^{\frac{1}{2}}\right]\right)^2.$$

Lemma 46. *For $A, B \in \mathcal{L}(\mathcal{K})$, $A, B \geq 0$, it holds*

1. $F(A,B) = F(B,A)$,
2. $F(\lambda A, B) = \lambda F(A,B)$ *for all* $\lambda \geq 0$,
3. $F(|\psi\rangle\langle\psi|, B) = \langle\psi, B\psi\rangle$ *for all* $\psi \in \mathcal{K}$, *and*
4. $0 \leq F(A,B) \leq 1$ *for density matrices* A, B.

The following theorem is a very important structural assertion for the quantum fidelity. It rephrases the fidelity as resulting from an optimization of the overlaps of purifications of the states.

Theorem 47 (Uhlmanns Theorem)**.** *Let $\rho, \sigma \in \mathcal{S}(\mathcal{K})$. It holds*

$$F(\rho,\sigma) = \max\left\{|\langle\psi,\varphi\rangle|^2 : \psi \text{ is purification of } \rho,\ \varphi \text{ is purification of } \sigma\right\}. \tag{5.1}$$

Before we give a proof of Uhlmanns theorem, we state and prove two supporting lemmas. The first one is a variational characterization of the trace norm.

Lemma 48. *Let $A \in \mathcal{L}(\mathcal{H})$. It holds*

$$\|A\|_1 := \max\{|\mathrm{tr}AU| : U \in \mathcal{L}(\mathcal{H}) \text{ unitary}\}. \tag{5.2}$$

Proof. Let $A = U_0|A|$ be a polar decomposition of A (remember the definition $|A| = \sqrt{A^*A}$), and

$$|A| = \sum_{i=1}^{m} a_i |\varphi_i\rangle\langle\varphi_i|$$

be a spectral decomposition of $|A|$. Let U be any unitary on $\mathcal{H}$ and define $V := UU_0$ (which is also a unitary matrix!). We have

$$\begin{aligned}
|\mathrm{tr}(AU)| &= |\mathrm{tr}(|A|V)| \\
&= |\sum_{i=1}^{m} a_i \langle\varphi_i, V\varphi_i\rangle| \\
&\leq \sum_{i=1}^{m} a_i |\langle\varphi_i, V\varphi_i\rangle| \\
&\leq \sum_{i=1}^{m} a_i \\
&= \mathrm{tr}|A| \\
&= \|A\|_1.
\end{aligned}$$

The first inequality above is the triangle inequality. The second one follows from the fact, that V is a unitary. The last equality is by definition of the trace norm. Therefore the right hand side is dominated by the left hand side in (5.2). On the other hand, the reverse inequality also holds (5.2) is achieved. Indeed, it holds

$$|\mathrm{tr}(AU_0^*)| = \mathrm{tr}(|A|) = \|A\|_1.$$

□

Lemma 49. *Let $\{\varphi_i\}_{i=1}^d$ an orthonormal basis in $\mathcal{K}$, $\dim\mathcal{K} := d$. And let $\phi \in \mathcal{K}\otimes\mathcal{K}$ be defined by $\phi := \sum_{i=1}^d \varphi_i \otimes \varphi_i$. It holds*

$$(A\otimes \mathbb{1}_{\mathcal{K}})\phi = (\mathbb{1}_{\mathcal{K}}\otimes A^T)\phi$$

where A^T is the transpose of A with respect to $\{\varphi_i\}_{i=1}^d$ [1].

Proof. Let $A = \sum_{i,k=1}^d a_{ij}\,|\varphi_i\rangle\langle\varphi_j|$, (written as a linear combination in the $\{|\varphi_i\rangle\langle\varphi_j|\}_{i,j=1}^d$ basis.) Straightforward calculation then gives

$$\begin{aligned}
(A\otimes\mathbb{1}_{\mathcal{K}})|\phi\rangle &= \sum_{i=1}^d A|\varphi_i\rangle\otimes|\varphi_i\rangle \\
&= \sum_{i,k,l=1}^d a_{kl}\,|\varphi_k\rangle\langle\varphi_l||\varphi_i\rangle\otimes|\varphi_i\rangle \\
&= \sum_{k,i=1}^d a_{ki}\,|\varphi_k\rangle\otimes|\varphi_i\rangle \\
&= \sum_{k=1}^d \left(|\varphi_k\rangle\otimes\sum_{i=1}^d a_{ki}\,|\varphi_i\rangle\right) \\
&= \sum_{k=1}^d |\varphi_k\rangle\otimes\left(\sum_{l,i=1}^d a_{li}\,|\varphi_i\rangle\langle\varphi_l,\varphi_k\rangle\right) \\
&= \sum_{k=1}^d |\varphi_k\rangle\otimes A^T|\varphi_k\rangle \\
&= (\mathbb{1}_{\mathcal{K}}\otimes A^T)|\phi\rangle
\end{aligned}$$

□

Proof of Theorem 47. Let ψ,φ be unit vectors in $\mathcal{K}\otimes\mathcal{K}$, and assume, that ψ is state vector of a purification of ρ and φ is state vector of a purification of σ. Let

$$\psi \;=\; \sum_{k=1}^d \sqrt{\mu_k}\psi_k\otimes\gamma_k, \quad \text{and} \qquad \varphi \;=\; \sum_{l=1}^d \sqrt{\vartheta_l}\varphi_l\otimes\tau_l$$

be Schmidt decompositions of ψ resp. φ (some Schmidt coefficients may vanish). Let $U_1, U_2, V \in \mathcal{L}(\mathcal{K})$ be unitaries such that for each $k\in[d]$ the equalities

$$\varphi_k = V\psi_k, \quad \gamma_k = U_1\psi_k, \text{ and } \quad \tau_m = U_2\varphi_k$$

hold. Definition of the square root function, we have the eigenvalue equations

$$\sqrt{\mu_k}\psi_k = \rho^{\frac{1}{2}}\psi_k \qquad \sqrt{\vartheta_k}\varphi_k = \sigma^{\frac{1}{2}}\varphi_k \tag{5.3}$$

[1] Note that the transposed matrix depends on the chosen basis

It holds

$$\begin{aligned}\psi &= \sum_{k=1}^{d} \sqrt{\mu_k}\psi_k \otimes \gamma_k \\ &= (\rho^{\frac{1}{2}} \otimes \mathbb{1}_{\mathcal{K}}) \sum_{k=1}^{d} \psi_k \otimes \gamma_k \\ &= (\rho^{\frac{1}{2}} \otimes \mathbb{1}_{\mathcal{K}}) \sum_{k=1}^{d} \psi_k \otimes U_1\psi_k \\ &= (\rho^{\frac{1}{2}} \otimes U_1) \sum_{k=1}^{d} \psi_k \otimes \psi_k \\ &= (\rho^{\frac{1}{2}} U_1^T \otimes \mathbb{1}_{\mathcal{K}}) \sum_{k=1}^{d} \psi_k \otimes \psi_k. \end{aligned} \tag{5.4}$$

The last of the equalities above is by Lemma 49. By a very similar calculation as done above for ψ, also

$$\varphi = (\sigma^{\frac{1}{2}} U_2^T \otimes \mathbb{1}_{\mathcal{K}}) \sum_{k=1}^{d} \varphi_k \otimes \varphi_k$$

holds. We further calculate

$$\varphi = (\sigma^{\frac{1}{2}} U_2^T \otimes \mathbb{1}_{\mathcal{K}}) \sum_{k=1}^{d} V\psi_k \otimes V\psi_k \tag{5.5}$$

$$= (\sigma^{\frac{1}{2}} U_2^T VV^T \otimes \mathbb{1}_{\mathcal{K}}) \sum_{k=1}^{d} \psi_k \otimes \psi_k, \tag{5.6}$$

where we once more used Lemma 49 to obtain the last equality above. With (5.4) and (5.6), we calculate

$$\begin{aligned}\langle \psi, \varphi \rangle &= \sum_{k,l=1}^{d} \langle (\rho^{\frac{1}{2}} U_1^T \otimes \mathbb{1}_{\mathcal{K}})\psi_k \otimes \psi_k, (\sigma^{\frac{1}{2}} U_2^T VV^T \otimes \mathbb{1}_{\mathcal{K}})\psi_l \otimes \psi_l \rangle \\ &= \sum_{k,l=1}^{d} \langle \psi_k \otimes \psi_k, (\rho^{\frac{1}{2}} U_1^T \otimes \mathbb{1}_{\mathcal{K}})^* (\sigma^{\frac{1}{2}} U_2^T VV^T \otimes \mathbb{1}_{\mathcal{K}})\psi_l \otimes \psi_l \rangle \\ &= \sum_{k,l=1}^{d} \langle \psi_k \otimes \psi_k, (\overline{U}_1 \rho^{\frac{1}{2}} \sigma^{\frac{1}{2}} U_2^T VV^T \otimes \mathbb{1}_{\mathcal{K}})\psi_l \otimes \psi_l \rangle \\ &= \sum_{k,l=1}^{d} \langle \psi_k, \overline{U}_1 \rho^{\frac{1}{2}} \sigma^{\frac{1}{2}} U_2^T VV^T \psi_l \rangle \cdot \langle \psi_k, \psi_l \rangle \\ &= \sum_{k=1}^{d} \langle \psi_k, \overline{U}_1 \rho^{\frac{1}{2}} \sigma^{\frac{1}{2}} U_2^T VV^T \psi_k \rangle \\ &= \mathrm{tr}(\overline{U}_1 \rho^{\frac{1}{2}} \sigma^{\frac{1}{2}} U_2^T VV^T). \end{aligned}$$

The above chain of equations yields

$$|\langle\psi,\varphi\rangle| \;=\; |\mathrm{tr}(\rho^{\frac{1}{2}}\sigma^{\frac{1}{2}}U_2^T V V^T \overline{U}_1)| \tag{5.7}$$

We note, that $U_2^T V V^T \overline{U}_1$ is a unitary matrix. Furthermore, U_1, U_2, V depend on the choice of purifications of the state. In fact, each unitary on $\mathcal{K}$ can be realized by $U_2^T V V^T \overline{U}_1$ by choosing corresponding purifications. If we choose purifications, such that right hand side of Eq. (5.7) is maximal, we obtain using Lemma 48 the claim of the Theorem. □

Definition 50 (Entanglement fidelity)**.** *Let $\rho \in \mathcal{S}(\mathcal{K})$, $\mathcal{N} \in \mathcal{C}(\mathcal{K},\mathcal{K})$. The entanglement fidelity of $(\rho,\mathcal{N})$ is defined by*

$$F_e(\rho,\mathcal{N}) := F(|\psi\rangle\langle\psi|, \mathrm{id}_{\mathcal{K}} \otimes \mathcal{N}(|\psi\rangle\langle\psi|) = \langle\psi, \mathrm{id}_{\mathcal{K}} \otimes \mathcal{N}(|\psi\rangle\langle\psi|)\psi\rangle, \tag{5.8}$$

where ψ is the state vector of a purification of ρ.

The reader may ask, whether or not the entanglement fidelity is well-defined. This question is affirmatively answered by the statement of the following lemma. In fact, one consequence of the representation in terms of Kraus operators given below is, that the entanglement fidelity does not depend on the chosen purification of the argument.

Lemma 51. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{K},\mathcal{K})$ and $\mathcal{N}(x) = \sum_{k=1}^N A_k x A_k^*$ $(x \in \mathcal{L}(\mathcal{K}))$ a Kraus decomposition of $\mathcal{N}$. It holds for each $\rho \in \mathcal{S}(\mathcal{K})$*

$$F_e(\rho,\mathcal{N}) = \sum_{k=1}^N |\mathrm{tr}(A_k\rho)|^2 \tag{5.9}$$

Proof. Let ψ be the state vector of any purification of ρ with Schmidt decomposition $\psi = \sum_{i=1}^d \sqrt{\lambda_i}\ \gamma_i \otimes \vartheta_i$. Define $\tilde{\vartheta}_i := \sqrt{\lambda_i}\vartheta_i$ for each $i \in [d]$. It holds

$$\psi = \sum_{i=1}^d \gamma_i \otimes \tilde{\vartheta}_i, \quad \text{and} \quad \rho = \sum_{i=1}^d |\tilde{\vartheta}_i\rangle\langle\tilde{\vartheta}_i| \tag{5.10}$$

We calculate

$$\begin{aligned}\langle\psi, \mathrm{id}_{\mathcal{K}} \otimes \mathcal{N}(|\psi\rangle\langle\psi|)\psi\rangle \;&=\; \sum_{i,j,l,m=1}^d \langle\gamma_i \otimes \tilde{\vartheta}_i, (\mathrm{id}_{\mathcal{K}} \otimes \mathcal{N})(|\gamma_j \otimes \tilde{\vartheta}_j\rangle\langle\gamma_l \otimes \tilde{\vartheta}_l|), \gamma_m \otimes \tilde{\vartheta}_m\rangle \\ &= \sum_{i,l=1}^d \langle\tilde{\vartheta}_i, \mathcal{N}(|\tilde{\vartheta}_i\rangle\langle\tilde{\vartheta}_l|), \tilde{\vartheta}_l\rangle. \end{aligned} \tag{5.11}$$

For each $i,l \in [d]$, the summand on the r.h.s. of the equality in (5.11) can be further written as

$$\langle\tilde{\vartheta}_i, \mathcal{N}(|\tilde{\vartheta}_i\rangle\langle\tilde{\vartheta}_l|), \tilde{\vartheta}_l\rangle \;=\; \sum_{k=1}^N \langle\tilde{\vartheta}_i, A_k|\tilde{\vartheta}_i\rangle\langle\tilde{\vartheta}_l|A_k^*, \tilde{\vartheta}_l\rangle \;=\; \sum_{k=1}^N \mathrm{tr}|\tilde{\vartheta}_i\rangle\langle\tilde{\vartheta}_i|A_k \cdot \overline{\mathrm{tr}|\tilde{\vartheta}_l\rangle\langle\tilde{\vartheta}_l|A_k}.$$

Combination of (5.11) with (5.12) leads us to

$$\begin{aligned}\langle\psi, \mathrm{id}_{\mathcal{K}} \otimes \mathcal{N}(|\psi\rangle\langle\psi|)\psi\rangle &= \sum_{k=1}^{N}\sum_{i,l=1}^{d} \mathrm{tr}|\tilde{\vartheta}_i\rangle\langle\tilde{\vartheta}_i|A_k \cdot \overline{\mathrm{tr}|\tilde{\vartheta}_l\rangle\langle\tilde{\vartheta}_l|A_k} \\ &= \sum_{k=1}^{N} \mathrm{tr}\sum_{i=1}^{d}|\tilde{\vartheta}_i\rangle\langle\tilde{\vartheta}_i|A_k \cdot \overline{\mathrm{tr}\sum_{l=1}^{d}|\tilde{\vartheta}_l\rangle\langle\tilde{\vartheta}_l|A_k} \\ &= \sum_{k=1}^{N}|\mathrm{tr}A_k\rho|^2.\end{aligned}$$

□

5.2 Quantum Source Compression

In this section, we will determine the optimal rate for compression of a DMQS. We define

Definition 52. *An* (n,k)-code for source compression *of the DMQS* $\rho \in \mathcal{S}(\mathcal{H})$ *is a pair* $(\mathcal{E},\mathcal{D})$, *where* $\mathcal{E} \in \mathcal{C}(\mathcal{H}^{\otimes n},\mathcal{K})$, $\mathcal{D} \in \mathcal{C}(\mathcal{K},\mathcal{H}^{\otimes n})$ *are c.p.t.p. maps, and* $k = \dim\mathcal{K}$. *We define for each* $n \in \mathbb{N}, \epsilon \geq 0$

$$K(\rho,n,\epsilon) := \min\left\{k : \exists(n,k)\text{-code } (\mathcal{E},\mathcal{D}) \text{ with } F_e(\rho^{\otimes n}, \mathcal{D}\circ\mathcal{E}) \geq 1-\epsilon\right\}$$

The following assertion is known as the source compression theorem for discrete memoryless quantum sources.

Theorem 53 (DMQS Source compression theorem)**.** *Let* $\rho \in \mathcal{S}(\mathcal{H})$. *It holds for all* $\epsilon \in (0,1)$

$$\lim_{n\to\infty}\frac{1}{n}\log K(\rho,n,\epsilon) = S(\rho).$$

The above theorem provides the von Neumann entropy S with an operational meaning. $S(\rho)$ is the minimal compression rate for asymptotically perfect compression of the DMQS ρ. Before we prove the theorem, we first recall some statements, we already have proven in Lemma 33.

Lemma 54. *Let* $\rho \in \mathcal{S}(\mathcal{H})$ *be a quantum state,* $\delta > 0$. *There exists a sequence* $\{p_n\}_{n=1}^{\infty}$ *of orthogonal projections such that for each* $n \in \mathbb{N}$

1. $[p_n, \rho^{\otimes n}] = 0$,
2. $2^{-n(S(\rho)+\delta))}p_n \leq p_n\rho^{\otimes n}p_n \leq 2^{-n(S(\rho)-\delta))}p_n$,
3. $\mathrm{tr}p_n \leq 2^{n(S(\rho)+\delta))}$, *and moreover*
4. $\lim_{n\to\infty}\mathrm{tr}p_n\rho^{\otimes n} = 1$

Proof. The claims follow from Lemma 33 with $\tau_1 = \tau_2 = 1$. The bound on the rank of p_n in 54.3 follows by

$$\mathrm{tr} p_n \leq 2^{n(S(\rho)+\delta)} \cdot \mathrm{tr} p_n \rho^{\otimes n} \ \leq \ 2^{n(S(\rho)+\delta)}.$$

□

We will first show the achievability part of Theorem 53.

Proposition 55. *Let $\rho \in \mathcal{S}(\mathcal{H})$. It holds*

$$\limsup_{n\to\infty} \frac{1}{n} \log K(\rho, n, \epsilon) \ \leq S(\rho) \tag{5.12}$$

for all $\epsilon > 0$.

Proof. Let $\epsilon \in (0,1)$, $\delta > 0$ be arbitrary but fixed numbers, and p_n for each $n \in \mathbb{N}$ the projection from Lemma 54. We show, that the inequality

$$K(\rho, n, \epsilon) \ \leq \ 2^{n(S(\rho)+\delta)} \tag{5.13}$$

holds for each large enough blocklength n (for all $n > n_0 := \min\left\{n \in \mathbb{N} : \ \mathrm{tr}(p_n \rho^{\otimes n}) > \sqrt{1-\epsilon}\right\}$). We define the shorthands $p_n^\perp := \mathbb{1}_{\mathcal{H}^{\otimes n}} - p_n$, and$\mathcal{K}_n := \mathrm{supp}(p_n)$. Fix a blocklength $n > n_0$. Define encoding and decoding channels $\mathcal{E}_n \in \mathcal{C}(\mathcal{H}^{\otimes n}, \mathcal{K})$, and $\mathcal{D}_n \in \mathcal{C}(\mathcal{K}_n, \mathcal{H}^{\otimes n})$ by

$$\begin{aligned} \mathcal{E}_n(A) &:= p_n A p_n^* + \mathrm{tr}(p_n^\perp) \frac{p_n}{\mathrm{tr}(p_n)} && (A \in \mathcal{L}(\mathcal{H}^{\otimes n})), \text{ and} \\ \mathcal{D}_n(B) &:= V B V^* && (B \in \mathcal{L}(\mathcal{K}_n)), \end{aligned}$$

where $V : \mathcal{K}_n \to \mathcal{H}^{\otimes n}$ is the injection of the subspace $\mathcal{K}_n$ into $\mathcal{H}^{\otimes n}$. Let $\mathcal{E}_n(\cdot) := \sum_{k=1}^N E_k(\cdot)E_k^*$ be a Kraus decomposition of $\mathcal{E}_n$ with $E_1 := p_n$ (check, that this is in fact possible). Using the corresponding Kraus decomposition

$$\mathcal{D}_n \circ \mathcal{E}_n(A) := \sum_{k=1}^N V E_k A (V E_k)^* \qquad (A \in \mathcal{L}(\mathcal{H}^{\otimes k}))$$

of the composition of encoding and decoding, $\mathcal{D}_n \circ \mathcal{E}_n$, we calculate

$$\begin{aligned} F_e(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) &= \sum_{k=1}^N |\mathrm{tr}(V E_k \rho^{\otimes n})|^2 && (5.14) \\ &\geq |\mathrm{tr}(V E_1 \rho^{\otimes n})|^2 \\ &= |\mathrm{tr}(p_n \rho^{\otimes n})|^2 \\ &\geq 1 - \epsilon. && (5.15) \end{aligned}$$

By the inequality in (5.15), we have

$$K(\rho, n, \epsilon) \ \leq \ \dim \mathcal{K}_n \ \leq \ \mathrm{tr}(p_n) \ \leq \ 2^{n(S(\rho)+\delta)}. \tag{5.16}$$

The rightmost inequality in (5.16) is from Lemma 54.3. We have shown the inequality in Eq. (5.13), which directly implies

$$\limsup_{n\to\infty} \frac{1}{n} \log K(\rho, n, \epsilon) \leq S(\rho) + \delta.$$

Since δ was an arbitrary positive number, we are done. □

Proposition 56. *Let $\rho \in \mathcal{S}(\mathcal{H})$. It holds for all $\epsilon \in (0,1)$*

$$\liminf_{n\to\infty} \frac{1}{n} \log K(\rho, n, \epsilon) \;\geq\; S(\rho).$$

Proof. Fix $\epsilon \in (0,1)$, and abbreviate $K_n := K(\rho, n, \epsilon)$ ($n \in \mathbb{N}$). Let $(\mathcal{E}_n, \mathcal{D}_n)$ be an (n, K_n) source compression code for ρ which fulfills

$$F_e(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) \geq 1 - \epsilon. \tag{5.17}$$

Let

$$\mathcal{E}_n(A) = \sum_{k=1}^{N} E_k A E_k^* \qquad (A \in \mathcal{L}(\mathcal{H}^{\otimes n})),$$
$$\mathcal{D}_n(B) = \sum_{l=1}^{M} D_l B D_l^* \qquad (B \in \mathcal{L}(\mathcal{K}_n))$$

be any Kraus decompositions of $\mathcal{E}_n$, $\mathcal{D}_n$. We represent the entanglement fidelity of the source compression protocol by means of its Kraus decompositions, we have, using Lemma 51,

$$F_e(\rho^{\otimes n}, \mathcal{D}_n \circ \mathcal{E}_n) = \sum_{k=1}^{N} \sum_{l=1}^{M} |\mathrm{tr}(D_l E_k \rho^{\otimes n})|^2. \tag{5.18}$$

We define the shorthands $\sigma := \rho^{\otimes n}$, and $\Delta_{kl} := |\mathrm{tr}(D_l E_k \sigma)|^2$ for each $l \in [N]$, $l \in [M]$, and denote by $\hat{q}_l$ to be the projector onto the range of D_l, in particular we have $\hat{q}_l D_l x = D_l x$ for each $x \in \mathcal{K}_n$. We estimate

$$\begin{aligned}
\Delta_{kl} &= |\mathrm{tr}(\sigma^{\frac{1}{2}} D_l E_k \sigma^{\frac{1}{2}})|^2 \\
&= |\mathrm{tr}(\sigma^{\frac{1}{2}} \hat{q}_l D_l E_k \sigma^{\frac{1}{2}})|^2 \\
&= |\langle \hat{q}_l \sigma^{\frac{1}{2}}, D_l E_k \sigma^{\frac{1}{2}} \rangle_{HS}|^2 \\
&\leq \|\hat{q}_l \sigma^{\frac{1}{2}}\|_2^2 \cdot \|D_l E_k \sigma^{\frac{1}{2}}\|_2^2 \\
&= \mathrm{tr}(\hat{q}_l \sigma) \cdot \mathrm{tr}(E_k^* D_l^* D_l E_k \sigma).
\end{aligned}$$

The inequality above is the Cauchy-Schartz inequality. Note, that with p_n (the projector from Lemma 54)

$$\mathrm{tr}(\hat{q}_l q) \;\leq\; \mathrm{tr}(\hat{q}_l) \;=\; \dim \mathrm{range} D_l \;\leq\; \dim \mathcal{K}_n \;=\; K_n$$

holds. Consequently

$$\mathrm{tr}(\hat{q}_l p_n \rho^{\otimes n} p_n) \;\leq\; 2^{-n(S(\rho)-\delta)} \mathrm{tr}(\hat{q}_l p_n) \;\leq\; 2^{-n(S(\rho)-\delta)} K_n.$$

We the shortcut $c_{kl} := \mathrm{tr}(E_k^* D_l^* D_l E_l \rho^{\otimes n})$ for each $k \in [N], l \in [M]$.

$$\begin{aligned}
1-\epsilon &\leq \sum_{k=1}^{N}\sum_{l=1}^{M} \mathrm{tr}(\hat{q}_l \rho^{\otimes n}) c_{kl} \\
&\leq \sum_{k=1}^{N}\sum_{l=1}^{M} \mathrm{tr}(\hat{q}_l (p_n + p_n^\perp))\rho^{\otimes n}(p_n + p_n^\perp)) c_{kl} \\
&= \sum_{k=1}^{N}\sum_{l=1}^{M} \Big(\mathrm{tr}(\hat{q}_l p_n \rho^{\otimes n} p_n) + \mathrm{tr}(\hat{q}_l p_n^\perp \rho^{\otimes n} p_n^\perp)\Big) c_{kl} \\
&\leq \sum_{k=1}^{N}\sum_{l=1}^{M} \Big(2^{-n(S(\rho)-\delta)} K_n + \mathrm{tr} p_n^\perp \rho^{\otimes n}\Big) c_{kl} \\
&= 2^{-n(S(\rho)-\delta)} K_n + \mathrm{tr} p_n^\perp \rho^{\otimes n} \qquad (5.19)
\end{aligned}$$

The last equation above is by the fact, that by definition $c_{kl} \geq 0$ and by properties of the Kraus decomposition $\sum_{k,l} c_{kl} = 1$. Rearranging the inequality in (5.19), we obtain

$$K(\rho, n, \epsilon) \ \geq (1 - \epsilon - \mathrm{tr} p_n^\perp \rho^{\otimes n}) \cdot 2^{n(S(\rho)-\delta)}.$$

We conclude

$$\liminf_{n\to\infty} \ \tfrac{1}{n} \log K(\rho, n, \epsilon) \ \geq S(\rho) - \delta.$$

Since $\delta > 0$ was arbitrary, we are done. □

We collect the obtained statements to prove the source compression theorem.

Proof of Theorem 53. It holds for each $\epsilon \in (0,1)$

$$\liminf_{n\to\infty} \tfrac{1}{n} \log K(\rho, n, \epsilon) \ \geq \ S(\rho) \ \geq \ \limsup_{n\to\infty} \tfrac{1}{n} \log K(\rho, n, \epsilon).$$

□

5.3 Supplement: Matrix Functions and some properties of the von Neumann entropy

We have seen, that the von Neummann entropy $S(\rho)$ of a density matrix ρ has an interpretation as the optimal asymptotical rate for compression of the memoryless quantum source generated by ρ. Here we state and prove some properties of this function.
First we show, that S is a concave function on $\mathcal{S}(\mathcal{H})$, i.e. the following statement.

Proposition 57. *Let $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$, and $\lambda \in (0,1)$. It holds*

$$S(\lambda\rho_1 + (1-\lambda)\rho_2) \ \geq \ \lambda S(\rho_1) + (1-\lambda) S(\rho_2). \qquad (5.20)$$

Proposition 58 (Peierl's Inequality)**.** *Let $f : I \to \mathbb{R}$ be convex, and $\{v_1, \ldots, v_n\} \subset \mathbb{C}^n$ be an orthonormal basis. It holds*

$$\sum_{j=1}^{n} f(\langle v_j, Av_j \rangle) \leq \operatorname{tr} f(A) \tag{5.21}$$

for each $A \in \mathcal{A}_n(I) := \{A \in \mathbb{M}_n : \operatorname{spec}(A) \subset I\}$.

Proof. Let

$$A = \sum_{j=1}^{m} a_j P_j \tag{5.22}$$

be spectral decomposition of A with $a_1, \ldots a_m \in I$ eigenvalues and P_j being the projection onto the eigenspace corresponding to a_j for each $j \in [m]$. By spectral calculus, $f(A)$ has spectral decomposition

$$A = \sum_{j=1}^{m} f(a_j) P_j. \tag{5.23}$$

Let $v \in \mathbb{C}^n$ be any unit vector. It holds

$$\begin{aligned} \langle v, f(A)v \rangle &= \sum_{j=1}^{m} f(a_j) \langle v, P_j v \rangle && (5.24) \\ &\geq f(\sum_{i=1}^{m} a_j \langle v, P_j v \rangle) && (5.25) \\ &= f(\langle v, Av \rangle). && (5.26) \end{aligned}$$

The inequality above follows from convexity of f together with the fact, that the term on the left-hand side is actually a convex combination of function values. The inequality then is by Jensen's inequality. We can apply the above inequality on all of the vectors $v_1, \ldots, v_n$ and obtain

$$\operatorname{tr} f(A) = \sum_{i=1}^{m} \langle v_i, f(A) v_i \rangle \geq \sum_{i=1}^{m} f(\langle v_i, Av_i \rangle). \tag{5.27}$$

□

Proposition 59. *If $f : I \to \mathbb{R}$ is convex, then the corresponding trace functional $\operatorname{tr} f : \mathcal{A}_n(I) \to \mathcal{L}(\mathcal{H})^h$ is convex.*

Proof. Fix $\lambda \in (0,1)$, and $A, B \in \mathcal{A}_n(I)$, and let $\{v_1, \ldots, v_i\}$ be an orthonormal basis of eigenvectors to

$$\lambda A + (1 - \lambda) B \tag{5.28}$$

(denote the eigenvalue corresponding to v_i by μ_i.) Then, it holds

$$\begin{aligned}
\mathrm{tr} f(\lambda A + (1-\lambda)B) &= \sum_{j=1}^{n} f(\mu_j) && (5.29)\\
&= \sum_{j=1}^{n} f(\langle v_j, (\lambda A + (1-\lambda)B) v_j \rangle) && (5.30)\\
&= \sum_{j=1}^{n} f(\lambda \langle v_j, \lambda A v_j \rangle + (1-\lambda)\langle v_j, B v_j \rangle) && (5.31)\\
&\leq \lambda \sum_{j=1}^{n} f(\langle v_j, A v_j \rangle) + (1-\lambda) \sum_{j=1}^{n} f(\langle v_j, B v_j \rangle) && (5.32)\\
&\leq \lambda \mathrm{tr} f(A) + (1-\lambda)\mathrm{tr} f(B). && (5.33)
\end{aligned}$$

The first inequality above is by convexity of f (as a scalar function), the second is by Peierl's inequality applied. □

Proof of Proposition 57. Proof is by application of Proposition 59. Note, that with $f : [0,1] \to \mathbb{R}$

$$f(x) := \begin{cases} x \cdot \log x & \text{if} x \in (0,1) \\ 0 & \text{if} x = 0, \end{cases} \tag{5.34}$$

it holds

$$S(\rho) = -\mathrm{tr} f(\rho). \tag{5.35}$$

Since f is convex on $[0,1]$, Proposition 59 tells us, that $-\mathrm{tr} f$ (and equivalently S) is concave. □

Proposition 60 (Monotonicity of S under pinching channels)**.** *Let $P_1, \ldots, P_K \subset \mathcal{L}(\mathcal{K})$ be mutually orthogonal projections (i.e. $P_k^* = P_k, P_k P_k = P_k$, and $P_k P_{k'} = 0, k' \neq k$. It holds*

$$S(\sum_{k=1}^{K} P_k \rho P_k) \geq S(\rho) \tag{5.36}$$

for all $\rho \in \mathcal{S}(\mathcal{K})$.

Proposition 61 (Almost-convexity of the von Neumann entropy)**.** *Let $\{\rho_x\}_{x\in\mathcal{X}} \subset \mathcal{S}(\mathcal{K})$ be a finite family of density matrices, and $q \in \mathcal{P}(\mathcal{X})$ a probability distribution. Define*

$$\overline{\rho} := \sum_{x\in\mathcal{X}} q(x)\rho_x. \tag{5.37}$$

It holds

1. $S(\overline{\rho}) \leq \sum_{x\in\mathcal{X}} q(x) S(\rho_x) + H(q)$.

2. *Equality in 1. holds if and only if* $\mathrm{supp}\rho_x \perp \mathrm{supp}\rho_{x'}$ *for all* $x \neq x'$.

Proof. We first prove the assertion in case of mixture of pure states. Let $\Psi_1, \ldots, \Psi_M \in \mathcal{S}(\mathcal{H})$ be pure states and $\mu \in \mathcal{P}([M])$. We show

$$S\left(\sum_{i=1}^{M} \mu(i)\ \Psi_i\right) \leq H(\mu). \tag{5.38}$$

Define, with an additional Hilbert space $\mathcal{H}' \simeq \mathbb{C}^M$, and an orthonormal basis $\{v_i\}_{i=1}^M$

$$\psi := \sum_{i=1}^{M} \sqrt{\mu(i)}\ \psi_i \otimes v_i \tag{5.39}$$

Notice, that $\mathrm{tr}_{\mathcal{H}'}|\psi\rangle\langle\psi| = \sum_{i=1}^M \mu(i)\Psi_i := \overline{\sigma}$. With $\Psi := |\psi\rangle\langle\psi|$, it also holds

$$S(\mathrm{tr}_{\mathcal{H}}\Psi) \ = \ S(\mathrm{tr}_{\mathcal{H}'}\Psi) = S(\overline{\sigma}). \tag{5.40}$$

Moreover, we have

$$\mathrm{tr}_{\mathcal{H}}\Psi \ = \ \sum_{i,j=1}^{M} \sqrt{\mu(i)\mu(j)}\langle\psi_i, \psi_j\rangle\ |v_i\rangle\langle v_j|. \tag{5.41}$$

If we now apply the pinching channel $\mathcal{P}(\cdot) := \sum_{k=1}^M |v_k\rangle\langle v_k|(\cdot)|v_k\rangle\langle v_k|$ on $\mathcal{H}'$, we obtain

$$\mathcal{P}(\mathrm{tr}_{\mathcal{H}}\Psi) \ = \ \sum_{i=1}^{M} \mu(i)|v_i\rangle\langle v_i| \tag{5.42}$$

which yields

$$S(\mathcal{P}(\mathrm{tr}_{\mathcal{H}}\Psi)) \ = \ H(\mu). \tag{5.43}$$

Putting everything together, we have

$$S(\overline{\sigma}) \ \leq \ S(\mathcal{P}(\mathrm{tr}_{\mathcal{H}}(\Psi)) \ = \ H(\mu). \tag{5.44}$$

Note that the inequality above is by Proposition 60. The general mixed-state case now easily follows. Let for each $i \in [N]$

$$\rho_i \ = \ \sum_{i=1}^{M_i} \mu_i(j)|v_j^{(i)}\rangle\langle v_j^{(i)}| \tag{5.45}$$

be a spectral decomposition of ρ_i. We then have

$$S\left(\sum_{i=1}^{N} \lambda(i)\rho_i\right) \ = \ S\left(\sum_{i=1}^{N}\sum_{j=1}^{M_i} \lambda(i)\mu_i(j)|v_j^{(i)}\rangle\langle v_j^{(i)}|\right) \tag{5.46}$$

$$\leq \ -\sum_{i=1}^{N}\sum_{j=1}^{M_i} \lambda(i)\mu_i(j)\log\lambda(i)\mu_i(j) \tag{5.47}$$

$$= H(\lambda) - \sum_{i=1}^{N} \lambda(i)S(\rho_i). \tag{5.48}$$

The inequality above is by applying the pure-state case. Rearrangement of the above inequalty proves the claim. Inspecting the inequality in (5.47) we also verify the second claim of the proposition. Therein equality holds, if all projections appearing in the sum are mutually orthogonal. □

Lemma 62. *Let $v : \mathcal{H} \to \mathcal{K}$ an isometric linear map (i.e. $vv^* = \mathbb{1}_{\mathcal{H}}$, and v^*v is an orthogonal projection in $\mathcal{K}$. Then*

$$S(v\rho v^*) \;=\; S(\rho) \tag{5.49}$$

does hold for all $\rho \in \mathcal{S}(\mathcal{H})$. In particular S is invariant under unitaries.

Proof. Is clear. Isometries do not change the spectrum of a matrix including multiplicities. □

Lemma 63. *Let $\rho \in \mathcal{S}(\mathcal{H})$. It holds*

$$0 \leq S(\rho) \leq \log \dim \mathcal{H}. \tag{5.50}$$

Proof. The bounds on $S(\rho)$ directly carry over from the Shannon entropy via spectral decomposition. We give a proof for the right hand inequality for convenience. Let $\{\lambda(x)\}_{x=1}^d$ the probability distribution arising from the spectrum of ρ (counting multiplicities). Let f be the function defined in (5.35). It holds

$$S(\rho) \;=\; H(\lambda) \;=\; -\sum_{x=1}^{d} f(x) \;= -d \cdot \sum_{x=1}^{d} \frac{1}{d} f(\lambda(x)) \;\leq\; -d \cdot f\left(\sum_{x=1}^{d} \frac{\lambda(x)}{d}\right) \;=\; -d \cdot f\left(\frac{1}{d}\right) = \log d.$$

The inequality above is by concavity of f. □

Lemma 64. *Let $\rho \in \mathcal{S}(\mathcal{H} \otimes \mathcal{K}), \sigma_1 := \mathrm{tr}_{\mathcal{K}}(\rho), \sigma_2 := \mathrm{tr}_{\mathcal{H}}(\rho)$. It holds*

$$S(\rho) \leq S(\sigma_1) + S(\sigma_2), \tag{5.51}$$

where equality holds if $\rho = \sigma_1 \otimes \sigma_2$.

Proof. We calculate

$$\begin{aligned}
\mathrm{tr}\rho \log(\sigma_1 \otimes \sigma_2) \;&=\; \mathrm{tr}\rho\,(\log \sigma_1 \otimes \mathbb{1}_{\mathcal{K}} + \mathbb{1}_{\mathcal{H}} \otimes \log \sigma_2) && (5.52)\\
&=\; \mathrm{tr}\rho(\log \sigma_1 \otimes \mathbb{1}_{\mathcal{K}}) \;+ \mathrm{tr}\rho \log(\mathbb{1}_{\mathcal{H}} \otimes \sigma_2) && (5.53)\\
&=\; \mathrm{tr}\sigma_1 \log \sigma_1 \;+\; \mathrm{tr}\sigma_2 \log \sigma_2 && (5.54)\\
&=\; -S(\sigma_1) \;-\; S(\sigma_2). && (5.55)
\end{aligned}$$

Then we have

$$\begin{aligned}
0 \;\leq D(\rho \| \sigma_1 \otimes \sigma_2) \;&=\; -S(\rho) + \mathrm{tr}(\rho \log(\sigma_1 \otimes \sigma_2)) && (5.56)\\
&=\; -S(\rho) + S(\sigma_1) + S(\sigma_2). && (5.57)
\end{aligned}$$

Rearrangement yields the desired inequality. □

5.4 Exercises

Exercise 65 (Further properties of the fidelity)**.** *Prove the following properties of F.*

1. ***Multiplicativity:*** $F(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2) = F(\rho_1,\sigma_1)\cdot F(\rho_2,\sigma_2)$ *for states* $\rho_1,\sigma_1\in\mathcal{S}(\mathcal{H})$, *and* $\rho_2,\sigma_2\in\mathcal{S}(\mathcal{K})$ *with some Hilbert spaces* $\mathcal{H},\mathcal{K}$.
2. $F(\rho,\sigma)=0 \Rightarrow \mathrm{supp}\rho\perp\mathrm{supp}\sigma$.
3. $F(\rho,\sigma)=1 \Rightarrow \rho=\sigma$
4. $F(\mathrm{tr}_2\rho,\mathrm{tr}_2\sigma) \geq F(\rho,\sigma)$ *for states* $\rho,\sigma\in\mathcal{S}(\mathcal{H}_1\otimes\mathcal{H}_2)$.
5. ***Isometric invariance:*** $F(V\rho V^*,V\sigma V^*)$ *for a linear isometry* $V:\ \mathcal{H}\to\mathcal{K}$.
6. ***C.P.T.P. Monotonicity:*** $F(\mathcal{T}(\rho),\mathcal{T}(\sigma))\geq F(\rho,\sigma)$ *for* $\mathcal{T}\in\mathcal{C}(\mathcal{H},\mathcal{K})$ *and* $\rho,\sigma\in\mathcal{S}(\mathcal{H})$.

Hints: For proving 1.-5. Uhlmann's Theorem might help. To show 6. Use properties 4. and 5. and remember Stinesprings dilation Theorem.

Exercise 66 (No Cloning)**.** *As we already know from the Lecture 4, the* universal quantum cloning device, *i.e. is an impossible machine, i.e. the map* $\mathcal{T}:\mathcal{L}(\mathcal{H})\to\mathcal{L}(\mathcal{H}^{\otimes 2})$ *with*

$$\mathcal{T}(\rho) = \rho\otimes\rho \qquad (\rho\in\mathcal{S}(\mathcal{H})) \tag{5.58}$$

is not a quantum channel.

1. *One could ask, whether or not a cloning channel can be found under the restriction, that it only has to succeed on some instead of all states in* $\mathcal{S}(\mathcal{H})$.
   Prove the following somewhat discouraging assertion.

   Theorem 67 (No cloning theorem)**.** *Let* $\rho_1,\rho_2\in\mathcal{S}(\mathcal{H})$ *be any two density matrices. There exists a c.p.t.p. map* $\mathcal{T}\in\mathcal{C}(\mathcal{H},\mathcal{H}\otimes\mathcal{H})$, *such that* $\mathcal{T}(\rho_i)=\rho_i\otimes\rho_i$ *holds for* $i=1,2$ *if and only if either* $\mathrm{supp}\rho_1\perp\mathrm{supp}\rho_2$ *or* $\rho_1=\rho_2$ *is true.*

   Hint: Use properties of the fidelity from Exercise 65 (This may be a two-line-proof).

2. *A somewhat less restrictive task is* broadcasting *of quantum states. A channel* $\mathcal{F}\in\mathcal{C}(\mathcal{H}_A,\mathcal{H}\otimes\mathcal{H}_B)$, $\mathcal{H}_A\simeq\mathcal{H}_B\simeq\mathcal{H}$ *is a broadcasting channel for a state* $\rho\in\mathcal{S}(\mathcal{H})$, *if and only if*

   $$\mathrm{tr}_A\mathcal{F}(\rho) = \mathrm{tr}_B\mathcal{F}(\rho) = \rho \tag{5.59}$$

   Let $\rho_1,\rho_2\in\mathcal{S}(\mathcal{H})$ *be commuting density matrices. Construct a channel, which is a broadcasting channel for both of them.*
   Hint: Use the fact, that commuting matrices are simultaneously diagonalizable and the Schmidt decomposition theorem.

Exercise 68. *Let* $\psi\in\mathcal{H}_A\otimes\mathcal{H}_B$ *be a unit vector, and* $\rho_A:=\mathrm{tr}_{\mathcal{H}_B}|\psi\rangle\langle\psi|$, $\rho_B:=\mathrm{tr}_{\mathcal{H}_A}|\psi\rangle\langle\psi|$ *the marginals of the corresponding pure state* $|\psi\rangle\langle\psi|$. *Show, that*

$$S(\rho_A)=S(\rho_B) \tag{5.60}$$

holds (Hint: Remember the Schmidt decomposition.)

6 Message transmission over quantum channels

In this lecture we devote ourselves to a discussion of classical message transmission over quantum channels. In Section 6.1, we discuss channel coding over a semiclassical model - channels with classical input and quantum output. We determine the message transmission of discrete memoryless channels of this type.
In Section 6.2, we generalize the model to a channel with quantum input and quantum output.

6.1 The discrete memoryless classical-quantum channel

In this section, we assume that sender and receiver are connected by a transmission line, where the input is a classical symbol while the output is a quantum system. This scenario is modeled by a so-called *classical-quantum channel* (or *cq channel*), which is a map

$$\begin{aligned} V:\ \mathcal{Y} &\to \mathcal{S}(\mathcal{K}), \\ y &\mapsto V(y) \in \mathcal{S}(\mathcal{K}). \end{aligned}$$

for some alphabet $\mathcal{Y}$ and Hilbert space $\mathcal{K}$. If many uses of such a channel are available in a way that the transmissions are all mutually independent, we model the transmission by the following memoryless channel model.

Definition 69 (Discrete memoryless classical-quantum channel)**.** *The* discrete memoryless classical quantum channel (DMCQC) *generated by a cq channel* $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ *is given by the family* $\{V^{\otimes n}\}_{n\in\mathbb{N}}$ *where for each* $n \in \mathbb{N}$ *the cq channel* $V^{\otimes n}$

$$V^{\otimes n}(x^n) := \bigotimes_{i=1}^{n} V(x_i) = V(x_1) \otimes \cdots \otimes V(x_n)$$

for each $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$.

Having defined a channel model, a standard task in information theory, is to give a general capacity formula which quantifies the message transmission abilities. We aim to determine the message transmission capacity of DMCQ channels defined above.
A channel code for n uses of a classical-quantum channel usually is given by a codeword u_m for each message m. Since the outputs of the channel are quantum mechanical systems, the decoding is performed by a quantum measurement (POVM) on the Hilbert space belonging to n outputs of channel (a general coding scheme is depicted in Fig. 6.1).

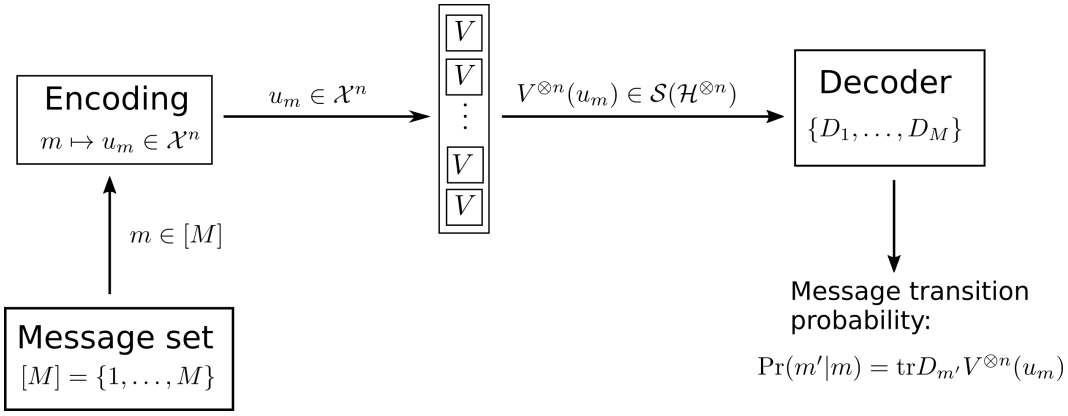


Figure 6.1: Coding scheme for classical message transmission over n uses of the DMCQC V.

We give rigorous definitions for the coding scenario.

Definition 70. *An (n, M)* code for classical message transmission *over the DMCQC generated by $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ is a family $\mathcal{C} := (u_m, D_m)_{m=1}^M$, where $u_1, \ldots, u_m$ are words in $\mathcal{X}^n$, and $\{D_m\}_{m=1}^M$ is a POVM in $\mathcal{L}(\mathcal{H}^{\otimes n})$. With the shortcut $D_m^c := \mathbb{1}_{\mathcal{H}}^{\otimes n} - D_m$, we define the functions*

$$\overline{e}(\mathcal{C}, V^{\otimes n}) := \frac{1}{M} \sum_{m=1}^{M} \mathrm{tr} D_m^c V^{\otimes n}(u_m) \qquad \textit{(average transmission error)},$$

$$e(\mathcal{C}, V^{\otimes n}) := \max_{m \in [M]} \mathrm{tr} D_m^c V^{\otimes n}(u_m) \qquad \textit{(maximal transmission error)}.$$

The quantities, which we aim to maximize are, for given transmission error $0 < \lambda < 1$, the maximal size of message sets for each blocklength n which allow transmission error being at most λ. We define the following quantities accordingly

$$\overline{N}(V, n, \lambda) := \max\{M \in \mathbb{N} : \exists\ (n, M) \text{ code } \mathcal{C} \text{ with } \overline{e}(\mathcal{C}, V^{\otimes n}) \leq \lambda\},$$
$$N(V, n, \lambda) := \max\{M \in \mathbb{N} : \exists\ (n, M) \text{ code } \mathcal{C} \text{ with } e(\mathcal{C}, W^{\otimes n}) \leq \lambda\}.$$

The quantities defined above grow in general exponentially with n (except, when the channel is completely useless for message transmission). We will determine the asymptotic behaviour of the *transmission rates*

$$\frac{1}{n} \log N(V, n, \lambda), \quad \text{and} \quad \frac{1}{n} \log \overline{N}(V, n, \lambda).$$

Exercise 71. *Show, that for each $n \in \mathbb{N}, \lambda \in (0, 1)$ and each cq channel V, the inequality*

$$N(V, n, \lambda) \leq \overline{N}(V, n, \lambda) \leq \frac{1}{1 - \sqrt{\lambda}} N(V, n, \sqrt{\lambda}).$$

holds. Hint: The left inequality follows directly from the definitions. The proof for the right inequality directly carries over from a corresponding relation for classical discrete memoryless channels.

For characterizing the message transmission capacity of the DMCQC we need the following function.

Definition 72 (Holevo quantity)**.** *Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a cq channel and $q \in \mathcal{P}(\mathcal{X})$ a probability distribution. The function*

$$\chi(q, V) := S(\overline{V}_q) - \sum_{x \in \mathcal{X}} q(x) S(V(x)) \tag{6.1}$$

with $\overline{V}_q := \sum_{x \in \mathcal{X}} q(x) V(x)$ is called the Holevo quantity *of (q, V).*

A convenient equivalent of the above expression can be given in terms of the quantum relative entropy. It holds by definition

$$\chi(q, V) = \sum_{x \in \mathcal{X}} q(x)\, D(V(x) \| \overline{V}_q). \tag{6.2}$$

Moreover, the *Holevo Information* of the channel V is defined by

$$C(V) \;:=\; \sup_{p \in \mathcal{P}(\mathcal{X})} \chi(p, V). \tag{6.3}$$

We will prove

Theorem 73 (Coding Theorem and Converse)**.** *Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a cq channel. The following statements are true.*

1. $\forall \lambda > 0 :\ \liminf_{n \to \infty} \frac{1}{n} \log N(V, n, \lambda) \;\geq\; C(V)$.
2. $\inf_{\lambda > 0}\ \limsup_{n \to \infty} \frac{1}{n} \log \overline{N}(V, n, \lambda) \;\leq\; C(V)$.

The first statement in the above theorem is usually called the *coding theorem* for the discrete memoryless classical-quantum channels, and the second claim the *(weak) converse* to the coding theorem.
Theorem 73 determines the *message transmission capacity* of a DMCQC V by $C(V)$. In fact, the second claim above can be replaced by the stronger statement

2'. $\forall \lambda > 0 :\ \limsup_{n \to \infty} \frac{1}{n} \log \overline{N}(V, n, \lambda) \leq \sup_{p \in \mathcal{P}(\mathcal{X})} \chi(p, V)$,

which is usually called the *strong converse* to the coding theorem for the DMCQC. The claim in 2'. does also hold, but we will not give a proof of this statement here.

The weak converse to the coding theorem

In this section, we aim to show the weak converse, i.e. the upper bound in Theorem 73.2. To show an instance of a *data processing inequality* for the Holevo quantity, we need

Theorem 74. *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$. It holds*

$$D(\rho \| \sigma) \;\geq\; D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)). \tag{6.4}$$

The monotonicity of the quantum relative entropy under c.p.t.p. maps (inequality in Eq. (6.4)) is provided with a very natural interpretation by Quantum Stein's Lemma (Theorem 30). The relative entropy $D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$ determines the optimal second kind error rate for quantum hypothesis tests with null hypothesis ρ and alternative hypothesis σ. This rate cannot be improved by affecting both hypotheses with "noise" (that is action of a c.p.t.p. map $\mathcal{N}$). This interpretation also is the idea behind the following "information-theoretic" proof of Theorem 74

Proof of Theorem 74. We assume that $\ker\sigma \subset \ker\rho$. Otherwise $D(\rho\|\sigma)$ is infinte which makes the inequality in (6.4) trivial. Since $\mathcal{N}$ is a linear map, it also holds $\ker\mathcal{N}(\rho) \subset \ker\mathcal{N}(\sigma)$ which ensures us, that both sides of the inequality in (6.4) are finite.
We first show the inequality

$$\beta_{\epsilon,n}(\rho,\sigma) \leq \beta_{\epsilon,n}(\mathcal{N}(\rho),\mathcal{N}(\sigma)) \tag{6.5}$$

for each $n \in \mathbb{N}, \epsilon > 0$. Let $\mathcal{N}_*$ be the Hilbert-Schmidt adjoint to $\mathcal{N}$ (remember Exercise 44). Because $\mathcal{N}_*$ is a completely positive and unital map, the image of the set

$$[0, \mathbb{1}_{\mathcal{K}}^{\otimes n}] := \{a \in \mathcal{L}(\mathcal{K})^{\otimes n} : \ 0 \leq a \leq \mathbb{1}_{\mathcal{K}}^{\otimes n}\}, \tag{6.6}$$

is contained in $[0, \mathbb{1}_{\mathcal{H}}^{\otimes n}]$. We can estimate

$$\begin{aligned}
\beta_{\epsilon,n}(\mathcal{N}(\rho),\mathcal{N}(\sigma)) &= \inf\left\{\mathrm{tr}(\mathcal{N}(\sigma)^{\otimes n}a) : \ a \in [0, \mathbb{1}_{\mathcal{K}}^{\otimes n}] \ \wedge \ \mathrm{tr}(\mathcal{N}(\rho)^{\otimes n}a) \geq 1-\epsilon\right\} \\
&= \inf\left\{\mathrm{tr}(\sigma^{\otimes n}\mathcal{N}_*^{\otimes n}(a)) : \ a \in [0, \mathbb{1}_{\mathcal{K}}^{\otimes n}] \ \wedge \ \mathrm{tr}(\rho^{\otimes n}\mathcal{N}_*^{\otimes n}(a)) \geq 1-\epsilon\right\} \\
&= \inf\left\{\mathrm{tr}(\sigma^{\otimes n}b) : \ b \in \mathcal{N}_*([0, \mathbb{1}_{\mathcal{K}}^{\otimes n}]) \ \wedge \ \mathrm{tr}(\rho^{\otimes n}b) \geq 1-\epsilon\right\} \\
&\geq \inf\left\{\mathrm{tr}(\sigma^{\otimes n}b) : \ b \in [0, \mathbb{1}_{\mathcal{H}}^{\otimes n}] \ \wedge \ \mathrm{tr}(\rho^{\otimes n}b) \geq 1-\epsilon\right\} \\
&= \beta_{\epsilon,n}(\rho,\sigma).
\end{aligned} \tag{6.7}$$

We conclude

$$\begin{aligned}
D(\rho\|\sigma) &= -\lim_{n\to\infty}\frac{1}{n}\log\beta_{\epsilon,n}(\rho,\sigma) \\
&\geq -\lim_{n\to\infty}\frac{1}{n}\log\beta_{\epsilon,n}(\mathcal{N}(\rho),\mathcal{N}(\sigma)) \\
&= D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)).
\end{aligned}$$

The inequality above is by (6.7). Both equalites are by using Quantum Stein's Lemma. □

Before we proceed, we recall some notation from classical information theory. For a stochastic matrix $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and probability distribution $p \in \mathcal{P}(\mathcal{X})$, the *mutual information* is defined by

$$I(p,W) \ := \ H(pW) - H(W|p),$$

where $pW \in \mathcal{P}(\mathcal{Y})$ is defined by $pW(y) := \sum_{x\in\mathcal{X}} p(x)W(y|x)$, and

$$H(W|p) := \sum_{x\in\mathcal{X}} p(x)H(W(\cdot|x)).$$

If the channel input and output in are regarded as a pair of random variables (X, Y) with distribution $P_{XY}(x, y) = p(x)W(y|x)$, also the notation

$$I(X \wedge Y) := H(X) - H(X|Y) \tag{6.8}$$

Based on the classical mutual information, we define for each classical-quantum channel $V : \mathcal{X} \to \mathcal{S}(\mathcal{K})$ and probability distribution $p \in \mathcal{P}(\mathcal{X})$ the so-called *accessible information*

$$\vartheta(p, V) := \sup\left\{I(p, \tilde{V}_{\mathcal{D}}) : \ \mathcal{D} \text{ is a (finite-valued) POVM on } \mathcal{K}\right\}$$

where $\tilde{V}_{\mathcal{D}}$ is defined as the effective $|\mathcal{X}|\times|\mathcal{Y}|$ stochastic matrix with the entries being the conditional measurement results if the POVM $\mathcal{D} := \{D_y\}_{y\in\mathcal{Y}}$ is applied on the channel ouputs, i.e.

$$\tilde{V}_{\mathcal{D}}(y|x) := \mathrm{tr} D_y V(x) \qquad\qquad (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{6.9}$$

The following proposition provides us with data processing inequalities we need for the converse proof.

Proposition 75. *Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{K})$, $p \in \mathcal{P}(\mathcal{X})$. The following claims hold.*

1. *For each quantum channel $\mathcal{N} \in \mathcal{C}(\mathcal{K}, \mathcal{H})$, it holds*

$$\chi(p, V) \geq \chi(p, \mathcal{N} \circ V). \qquad \textit{(Data processing inequality)}$$

2. *In particular, it holds*

$$\chi(p, V) \geq \vartheta(p, V). \qquad \textit{(Holevo bound)}$$

Proof. To prove the first claim, we consider the cq channel $\mathcal{N} \circ V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ which arises from the concatenation of V and $\mathcal{N}$, i.e.

$$\mathcal{N} \circ V(x) = \mathcal{N}(V(x)) \qquad\qquad (x \in \mathcal{X}).$$

Using the alternative formula (6.2) for the Holevo quantity, we have

$$\begin{aligned}
\chi(p, V) &= \sum_{x\in\mathcal{X}} p(x)\, D(V(x)\|\overline{V}_p) \\
&\geq \sum_{x\in\mathcal{X}} p(x)\, D(\mathcal{N}(V(x))\|\mathcal{N}(\overline{V}_p)) \\
&= \sum_{x\in\mathcal{X}} p(x)\, D(\mathcal{N}(V(x))\|\overline{(\mathcal{N} \circ V)}_p) \\
&= \chi(p, \mathcal{N} \circ V).
\end{aligned}$$

The first and the last equality above are by (6.2) together with linearity of $\mathcal{N}$ (i.e. $\mathcal{N}(\overline{V}_p) = \overline{(\mathcal{N} \circ V)}_p$). The inequality is by monoticity of the quantum relative entropy under c.p.t.p. maps (Theorem 74).

It remains to prove the second claim. Let $\mathcal{D} := \{D_y\}_{y\in\mathcal{Y}} \subset \mathcal{L}(\mathcal{K})$ be any POVM with $|\mathcal{Y}| < \infty$. We show the inequality

$$\chi(p, V) \geq I(p, \tilde{V}_{\mathcal{D}}). \tag{6.10}$$

Define a c.p.t.p. map $\hat{\mathcal{D}} \in \mathcal{C}(\mathcal{K}, \mathbb{C}^{|\mathcal{Y}|})$ by

$$\hat{\mathcal{D}}(a) := \sum_{y\in\mathcal{Y}} \operatorname{tr}(D_y a)\ |e_y\rangle\langle e_y| \qquad\qquad (a \in \mathcal{L}(\mathcal{K})).$$

It holds

$$\begin{aligned}\hat{\mathcal{D}} \circ V(x) \ &= \ \sum_{y\in\mathcal{Y}} \operatorname{tr}(D_y V(x))\ |e_y\rangle\langle e_y| \\ &= \ \sum_{y\in\mathcal{Y}} \tilde{V}_{\mathcal{D}}(y|x)\ |e_y\rangle\langle e_y|.\end{aligned}$$

Therefore, it holds (remember the definition in (6.9))

$$\begin{aligned}\chi(p, V) \ &\geq \ \chi(p, \hat{\mathcal{D}} \circ V) \\ &= \ S\left(\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} \tilde{V}_{\mathcal{D}}(y|x)\ |e_y\rangle\langle e_y|\right) - \sum_{x\in\mathcal{X}} p(x) S\left(\sum_{y\in\mathcal{Y}} \tilde{V}_{\mathcal{D}}(y|x)\ |e_y\rangle\langle e_y|\right) \\ &= \ H(p \circ \tilde{V}_{\mathcal{D}}) - H(\tilde{V}_{\mathcal{D}}|p) \qquad (6.11) \\ &= \ I(p, \tilde{V}_{\mathcal{D}}).\end{aligned}$$

Note, that the equality in (6.11) is true, because the arguments of the von Neumann entropies in the preceding line are in fact in diagonal form. Maximizing over both sides of the above inequality proves the claim, we obtain

$$\chi(p, V) \ \geq \ \sup\{I(p, \tilde{V}_{\mathcal{E}}) : \ \mathcal{E} \text{ is a finite POVM on } \mathcal{H}\} = \vartheta(p, V).$$

□

The inequalities shown in the preceding proposition allow proof of some additivity properties of the Holevo quantity.

Proposition 76. *Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{K})$, $W : \mathcal{Y} \to \mathcal{S}(\mathcal{H})$ be classical-quantum channels. The following claims are true*

1. $\sup_{p\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})} \chi(p, V \otimes W) \ = \ \sup_{q\in\mathcal{P}(\mathcal{X})} \chi(q, V) + \sup_{r\in\mathcal{P}(\mathcal{Y})} \chi(r, W).$

2. *For each $n \in \mathbb{N}$, it holds*

$$\frac{1}{n} \sup_{p\ \in\mathcal{P}(\mathcal{X}^n)} \chi(p, V^{\otimes n}) \ = \ \sup_{q\in\mathcal{P}(\mathcal{X})} \chi(q, V).$$

Proof. We first show the "≥" inequality in 1. For fixed probability distributions $q \in \mathcal{P}(\mathcal{X})$, $r \in \mathcal{P}(\mathcal{Y})$, it holds with $q \otimes r$ being the notation for the product distributions to q, r

$$\begin{aligned}\chi(q,V)+\chi(r,W) &= S(\overline{V}_q)+S(\overline{W}_r)-\sum_{x\in\mathcal{X}} q(x)S(V(x))-\sum_{y\in\mathcal{Y}} r(y)\ S(V(y)) \\ &= S\left(\overline{V}_q\otimes\overline{W}_r\right)-\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} q\otimes r(x,y)\cdot(S(V(x))+S(W(y))) \qquad (6.12)\\ &= S\left(\overline{(V\otimes W)}_{q\otimes r}\right)-\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} q\otimes r(x,y)S(V\otimes W(x,y)) \\ &= \chi(q\otimes r, V\otimes W). \qquad (6.13)\end{aligned}$$

The equality in (6.12) above is by additivity of the von Neumann entropy for product states (i.e. the case of equality in Lemma 64). Consequently, we have

$$\begin{aligned}\sup_{p\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})}\chi(p,V\otimes W) &\geq \sup_{q\in\mathcal{P}(\mathcal{X})}\sup_{r\in\mathcal{P}(\mathcal{Y})}\chi(q\otimes r,V\otimes W) \\ &= \sup_{q\in\mathcal{P}(\mathcal{X})}\chi(q,V)+\sup_{r\in\mathcal{P}(\mathcal{Y})}\chi(r,W). \qquad (6.14)\end{aligned}$$

The inequality above is by restricting the set for maximization. The equality follows by maximizing over the equality in (6.13).
To show the reverse inequality, fix a probability distribution $p \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$. We denote the marginal distributions on $\mathcal{X}$ and $\mathcal{Y}$ deriving from p by q and r, i.e. we set

$$q(x) := \sum_{y\in\mathcal{Y}} p(x,y) \quad \text{and} \quad r(y) := \sum_{x\in\mathcal{X}} p(x,y)$$

for each $x \in \mathcal{X}, y \in \mathcal{Y}$. It then holds

$$\mathrm{tr}_{\mathcal{H}}\overline{(V\otimes W)}_p = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\ \mathrm{tr}_{\mathcal{H}}(V(x)\otimes W(y)) = \sum_{x\in\mathcal{X}} q(x)\ U(x) = \overline{V}_q.$$

By an analogous calculation, the equality

$$\mathrm{tr}_{\mathcal{K}}\overline{(V\otimes W)}_p = \overline{W}_r$$

becomes clear. By the above equalities and subadditivity of the von Neumann entropy (Lemma 64), it holds

$$S\left(\overline{(V\otimes W)}_p\right) \leq S(\overline{V}_q)+S(\overline{W}_r). \qquad (6.15)$$

Moreover, we have by additivity of the von Neumann entropy for tensor product states

$$\begin{aligned}\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\ S(V(x)\otimes W(y)) &= \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\ S(V(x))+\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\ S(V(y)) \\ &= \sum_{x\in\mathcal{X}} q(x)\ S(V(x))+\sum_{y\mathcal{Y}} r(y)\ S(V(y)). \qquad (6.16)\end{aligned}$$

Thus

$$\begin{aligned}\chi(p, V\otimes W) &= S\left(\overline{(V\otimes W)_p}\right) - \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\, S(V(x)\otimes W(y)) \\ &\underset{(6.15),(6.16)}{\leq} S(\overline{V}_q) + S(\overline{W}_r) - \sum_{x\in\mathcal{X}} q(x) S(V(x)) - \sum_{y\in\mathcal{Y}} r(y) S(W(y)) \\ &= \chi(q, V) + \chi(r, W). \end{aligned} \tag{6.17}$$

Maximisation over the input probability distribution and the corresponding marginals on both sides of the inequality in (6.17) leads us to

$$\sup_{p\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})} \chi(p, V\otimes W) \leq \sup_{q\in\mathcal{P}(\mathcal{X})} \sup_{r\in\mathcal{P}(\mathcal{Y})} \chi(q\otimes r, V\otimes W) = \sup_{q\in\mathcal{P}(\mathcal{X})} \chi(q, V) + \sup_{r\in\mathcal{P}(\mathcal{Y})} \chi(r, W),$$

which shows, combined with (6.17) the first claim of the proposition.
The second claim is readily verified by iteratively using the first claim, i.e.

$$\begin{aligned}\sup_{p\in\mathcal{P}(\mathcal{X}^n)} \chi(p, V^{\otimes n}) &= \sup_{q\in\mathcal{P}(\mathcal{X})} \chi(p, V) + \sup_{r\in\mathcal{P}(\mathcal{X}^{(n-1)})} \chi(p, V^{\otimes(n-1)}) \\ &= 2\cdot \sup_{q\in\mathcal{P}(\mathcal{X})} \chi(p, V) + \sup_{r\in\mathcal{P}(\mathcal{X}^{(n-2)})} \chi(p, V^{\otimes(n-2)}) \\ &\vdots \\ &= n\cdot \sup_{q\in\mathcal{P}(\mathcal{X})} \chi(p, V).\end{aligned}$$

□

The next lemma is well-known in classical information theory. The proof can be found in any textbook on (classical) information theory.

Lemma 77 (Fano's Lemma)**.** *Let (X, X') be a pair of random variables with values in $\mathcal{X}$, and $\gamma := \Pr(X \neq X')$. It holds*

$$H(X'|X) \leq \gamma \log|\mathcal{X}| + h(\gamma). \tag{6.18}$$

with $h(t) := -t\log t - (1-t)\log(1-t)$ being the binary Shannon entropy for $t \in (0,1)$.

The following proposition directly implies the weak converse statement in Theorem 73

Proposition 78. *Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel. It holds*

$$\overline{N}(V, n, \lambda) \leq \exp\left(n\cdot C(V) + n\cdot\lambda\log|\mathcal{X}| + 1\right) \tag{6.19}$$

for each $n \in \mathbb{N}$, $\lambda \in (0,1)$.

Proof. Let $n \in \mathbb{N}$, $\lambda \in (0,1)$ be arbitrary but fixed. Let $\mathcal{C} := (u_m, D_m)_{m=1}^M$ be any (n, M)-code for classical message transmission over the DMCQC V which fulfills

$$\overline{e}(\mathcal{C}, V^{\otimes n}) \leq \lambda. \tag{6.20}$$

We show, that the right hand side of Eq. (6.19) is an upper bound on the number of messages in $\mathcal{C}$. Let $p_* \in \mathcal{P}(\mathcal{X}^n)$ be the equidistribution on the set of codewords of $\mathcal{C}$, i.e.

$$p_*(x^n) := \begin{cases} \frac{1}{M} & \text{if } x^n \text{ is a codeword in } \mathcal{C} \\ 0 & \text{otherwise} \end{cases}$$

It holds

$$\begin{aligned} n \cdot \sup_{p \in \mathcal{P}(\mathcal{X})} \chi(p, V) &= \sup_{q \in \mathcal{P}(\mathcal{X}^n)} \chi(q, V^{\otimes n}) \\ &\geq \chi(p_*, V^{\otimes n}) \\ &\geq \vartheta(p_*, V^{\otimes n}). \end{aligned} \tag{6.21}$$

The equality above is by Proposition 76.2, the last inequality follows from Holevo's bound Proposition 75.2. Using the decoding POVM $\mathcal{D} := \{D_m\}_{m=1}^M$ from $\mathcal{C}$, we define a stochastic matrix $\tilde{V}_{n,\mathcal{D}} : \mathcal{X}^n \to \mathcal{P}([M])$ defined by entries

$$\tilde{V}_{n,\mathcal{D}}(m|x^n) := \operatorname{tr} V^{\otimes n}(x^n) D_m \qquad (x^n \in \mathcal{X}^n, m \in [M]).$$

Using (6.21), we obtain

$$n \cdot \sup_{p \in \mathcal{P}(\mathcal{X})} \chi(p, V) \;\geq\; \vartheta(p_*, V^{\otimes n}) \;=\; \sup_{\mathcal{E} \text{ POVM}} I(p_*, (\widetilde{V^{\otimes n}})_{\mathcal{E}}) \;\geq\; I(p_*, \tilde{V}_{n,\mathcal{D}}). \tag{6.22}$$

Let (X, Y) be the pair of random variables defined by

$$\Pr(X = m \;\wedge\; Y = m') \;= \tilde{V}_{n,\mathcal{D}}(m'|x_m) p_*(x_m).$$

Then,

$$I(p_*, \tilde{V}_{n,\mathcal{D}}) \;=\; I(X \wedge Y) \underset{(6.8)}{=} H(X) - H(X|Y).$$

Because we chose p_* to be equidistributed, it holds $H(X) = H(p_*) = \log M$. Note, that also

$$\begin{aligned} \gamma := \Pr(X \neq Y) &= \sum_{m=1}^{M} \sum_{m' \neq m} \Pr(X = m \wedge Y = m') \\ &= \sum_{m=1}^{M} \sum_{m' \neq m} p^n(u_m) \cdot \operatorname{tr} D_{m'} V^{\otimes n}(u_m) \\ &= \frac{1}{M} \sum_{m=1}^{M} \operatorname{tr}(\mathbb{1} - D_m) V^{\otimes n}(u_m) \\ &= \overline{e}(\mathcal{C}, V^{\otimes n}) \\ &\leq \lambda \end{aligned} \tag{6.23}$$

holds. Using Fano's inequality (Lemma 77) in combination with the above equality, we further bound the right hand side of Eq. (6.22). In fact,

$$I(p_*, \tilde{V}_{n,\mathcal{D}}) \;=\; H(X) - H(X|Y) \;\geq \log M - \lambda \log M - h(\lambda).$$

We end up at

$$n \cdot \sup_{p \in \mathcal{P}(\mathcal{X})} \chi(p, V) \geq \log M - \lambda \log M - 1 \geq \log M - \lambda n \log |\mathcal{X}| - 1,$$

which can be rearranged to

$$\log M \leq nC(V) + n \cdot \lambda \log |\mathcal{X}| + 1.$$

Since $\mathcal{C}$ was an arbitrary code with average error bounded from above by λ, maximising M over all such codes yields the claimed inequality. □

The coding theorem

In this section, we aim to show the "achievability part" of Theorem 73. The statement follows directly from the claim of the following proposition.

Proposition 79. *Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a cq channel. For each $\lambda \in (0,1)$, and $\delta > 0$ there is a number $n_0 := n_0(\lambda, \delta)$, such that for all $n \geq n_0$ the inequality*

$$\overline{N}(W, n, \lambda) \geq \exp(n(C(V) - \delta))$$

is true.

One of the main "nonclassical" challenges in our proof of Proposition 79 is to replace a matrix product by a sum. The following matrix inequality provides a suitable replacement for the "union bound" which would be used in the corresponding classical proof.

Lemma 80 (Hayashi-Nagaoka inequality)**.** *Let $a, b \in \mathcal{L}(\mathcal{K})$ be matrices with $0 \leq a \leq \mathbb{1}_{\mathcal{K}}$, $b \geq 0$. It holds*

$$\mathbb{1}_{\mathcal{K}} - (a+b)^{-\frac{1}{2}} a (a+b)^{-\frac{1}{2}} \leq 2(\mathbb{1}_{\mathcal{K}} - a) + 4b, \tag{6.24}$$

where y^{-1} is the generalized inverse of y.

Proof. We can assume without loosing generality that $a+b$ has full rank. This can be seen as follows. Let P denote projector onto the support of $a+b$. Note, that P as well as $P^{\perp} := \mathbb{1} - P$ commute with $a, b, a+b$. The inequality in 6.24 holds therefore if and only if the inequalities

$$P - (a+b)^{-\frac{1}{2}} a (a+b)^{-\frac{1}{2}} P \leq 2P - PaP + 4PbP \tag{6.25}$$

and

$$P^{\perp} - P^{\perp}(a+b)^{-\frac{1}{2}} a (a+b)^{-\frac{1}{2}} P^{\perp} \leq 2P^{\perp} - P^{\perp} a P^{\perp} + 4P^{\perp} b P^{\perp} \tag{6.26}$$

simultaneously hold. The inequality in (6.26) is in fact the trivial inequality $P^{\perp} \leq 2P^{\perp}$. The inequality in (6.25) regarded as a matrix inequality on $\mathrm{supp}(a+b)$, i.e. a version of the original inequality in (6.24) where $(a+b)$ has full rank.
First, we note, that for given matrices x, y, w, z it holds

$$(x - \mathbb{1}) b (x - \mathbb{1}) = xbx - xb - bx + b,$$

which with some rearrangements leads us to

$$xbx \;=\; b + (x-\mathbb{1})b + b(x-\mathbb{1}) + (x-\mathbb{1})b(x-\mathbb{1}).$$

Moreover, it holds $(w-z)^*(w-z) \geq 0$, which implies

$$w^*z + z^*w \;\leq w^*w + z^*z.$$

which applied with $w := \sqrt{y}(x-\mathbb{1})$ and $z := \sqrt{y}$ reads

$$(x-\mathbb{1})b + b(x-\mathbb{1}) \;\leq\; (\mathbb{1}-x)b(\mathbb{1}-x) + b \tag{6.27}$$

We set $x := (a+b)^{-\frac{1}{2}}$. We have

$$\begin{aligned}
\mathbb{1} - (a+b)^{-\frac{1}{2}}a(a+b)^{-\frac{1}{2}} &= \mathbb{1} - xax \\
&= x(a+b)x - xax \\
&= xbx \\
&= b + w^*z + z^*w + (x-\mathbb{1})b(x-\mathbb{1}) \\
&\overset{(6.27)}{\leq} 2b + 2(x-\mathbb{1})b(x-\mathbb{1}) \\
&\overset{(*)}{\leq} 2b + 2(x-\mathbb{1})x^{-2}(x-\mathbb{1}) \\
&= 2b + 2(\mathbb{1} - 2x^{-1} + x^{-2}) \\
&= 2b + 2(\mathbb{1} - 2(a+b)^{\frac{1}{2}} + a + b) \\
&\overset{(**)}{\leq} 2b + 2(\mathbb{1} - 2a + a + b) \\
&= 4b + 2(\mathbb{1} - 2a)
\end{aligned}$$

The equalties above are by rearrangement of terms, $(*)$ is by the inequality $b \leq (a+b) = x^{-2}$, and $(**)$ is true, because the map $x \mapsto x^{\frac{1}{2}}$ is matrix monotone, therefore $a \leq a^{-\frac{1}{2}} \leq (a+b)^{-\frac{1}{2}}$. □

Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, and $q \in \mathcal{P}(\mathcal{X})$ be a probability distribution. We define density matrices $\rho(p,V)$ and $\sigma(p,V)$ on $\mathbb{C}^{|\mathcal{X}|} \otimes \mathcal{H}$ by

$$\rho(p,V) \;:=\; \sum_{x\in\mathcal{X}} q(x)|e_x\rangle\langle e_x| \otimes V(x) \tag{6.28}$$

$$\sigma(p,V) \;:=\; \sum_{x\in\mathcal{X}} q(x)|e_x\rangle\langle e_x| \otimes \overline{V}_q, \tag{6.29}$$

where we again used the definition $\overline{V}_q := \sum_{x\in\mathcal{X}} V(x)$.

Exercise 81. *Show, that for $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$, and $q \in \mathcal{P}(\mathcal{X})$*

$$\chi(q,V) \;=\; D(\rho(q,V)\|\sigma(q,V))$$

holds for the states defined in (6.28) and (6.29). Hint: Realise, that both matrices can be written as a Block-diagonal matrix, then use the definition of $D(\cdot\|\cdot)$.

Proposition 82 (Random coding)**.** *Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, and $q \in \mathcal{P}(\mathcal{X})$. Assume, we find for given $n \in \mathbb{N}$, $\lambda \in (0,1)$, $\tau > 0$ an effect $a \in \mathcal{L}(\mathcal{H} \otimes \mathbb{C}^{|\mathcal{X}|})^{\otimes n}$, $0 \leq a \leq \mathbb{1}^{\otimes n}_{\mathcal{H}\otimes\mathbb{C}^{|\mathcal{X}|}}$ which fulfills the conditions*

1. $\mathrm{tr} a\rho(q,V)^{\otimes n} \geq 1 - \lambda$ *and*
2. $\mathrm{tr} a\sigma(q,V)^{\otimes n} \leq \tau$.

Then there exists for each $M \in \mathbb{N}$ an (n,M)-code $\mathcal{C} = (u_m, D_m)_{m=1}^M$ with

$$\overline{e}(\mathcal{C}, V^{\otimes n}) \;<\; 2\lambda + 4M\tau.$$

Proof. Fix $n \in \mathbb{N}$. We set shortcuts $\rho := \rho(q,V)$, and $\sigma := \sigma(q,V)$. Let

$$v : \left(\mathbb{C}^{|\mathcal{X}|} \otimes \mathcal{H})^{\otimes n}\right)\otimes n \;\to\; (\mathbb{C}|\mathcal{X}|)^{\otimes n} \otimes \mathcal{H}^{\otimes n}$$

be the isometric linear matrix, which permutes the tensor factors in a way, that

$$v\left(\bigotimes_{i=1}^n (a_i \otimes b_i)\right) = \left(\bigotimes_{i=1}^n a_i\right) \otimes \left(\bigotimes_{i=1}^n b_i\right)$$

for all $a_1, \ldots, a_n \in \mathbb{C}|\mathcal{X}|$, $b_1, \ldots, b_n \in \mathcal{H}$. We will use permuted versions of ρ, and σ defined by

$$\begin{aligned}
\hat{\rho}_n &:= v\rho^{\otimes n}v^* = \sum_{x^n \in \mathcal{X}^n} q^n(x^n) |e_{x^n}\rangle\langle e_{x^n}| \otimes V^{\otimes n}(x^n) \\
\hat{\sigma}_n &:= v\sigma^{\otimes n}v^* = \sum_{x^n \in \mathcal{X}^n} q^n(x^n) |e_{x^n}\rangle\langle e_{x^n}| \otimes \overline{V}_q^{\otimes n}
\end{aligned}$$

with $\hat{a} := vav^*$, $\hat{a}$ can without loosing generality be assumed to be of the form

$$\hat{a} \;=\; \sum_{x^n \in \mathcal{X}^n} |e_{x^n}\rangle\langle e_{x^n}| \otimes h_{x^n} \tag{6.30}$$

with matrices $0 \leq h_{x^n} \leq \mathbb{1}^{\otimes n}_{\mathcal{H}}$ (why?). We have

$$\begin{aligned}
\mathrm{tr}\hat{a}\hat{\rho}_n &= \mathrm{tr} a\rho^{\otimes n} \geq 1 - \lambda, \text{ and} \\
\mathrm{tr}\hat{a}\hat{\sigma}_n &= \mathrm{tr} a\sigma^{\otimes n} \leq \tau.
\end{aligned}$$

Let, for a fixed number $M \in \mathbb{N}$, $U := (U_1, \ldots, U_M)$ be an independent and identically distributed family of random variables, each with values in $\mathcal{X}^n$ and

$$\Pr(U_1 = x^n) \;=\; q^n(x^n)$$

for all $x^n \in \mathcal{X}^n$. Each outcome $u = (u_1, \ldots, u_M)$ of U is assumed to be a family of codewords and we define an (n,M) code by introducing a decoding POVM setting

$$\tilde{h}(u) := \sum_{m=1}^M h_{u_m}$$

with h_{x^n} being the corresponding effect from (6.30) for each $x^n \in \mathcal{X}^n$. We define

$$D_m(u) := \tilde{h}(u)^{-\frac{1}{2}} h_{u_m} \tilde{h}^{-\frac{1}{2}}(u)$$

(note that we may possibly make use of the pseudoinverse in the above formula). The definitions above assure us, that $\{D_m\}_{m=1}^M$ is indeed a POVM (check this!), which makes $\mathcal{C}(u) := (u_m, D_m(u))_{m=1}^M$ a proper (n, M) classical message transmission code. We will now show, that the expectation of the random code fulfills

$$\mathbb{E}\overline{e}(\mathcal{C}, V^{\otimes n}) < 2\lambda + 4M\tau. \tag{6.31}$$

We apply Lemma 80 for fixed $u = (u_1, \ldots, u_M)$, and $m \in [M]$ with the correspondences

$$a \leftarrow h_{u_m} \text{ and } \quad b \leftarrow \tilde{h}(u) - h_{u_m} = \sum_{l \neq m} h_{u_l}.$$

The conclusion of Lemma 80 yields

$$D_m^c(u) \leq 2(\mathbb{1} - h_{u_m}) + 4 \cdot \sum_{l \neq m} h_{u_l}$$

Which via monotonicity of the trace under the Löwner partial order implies the bound

$$\mathrm{tr}\Big(D_m^c(u) V^{\otimes n}(u_m)\Big) \leq 2 \cdot \mathrm{tr}\Big((\mathbb{1} - h_{u_m}) W^{\otimes n}(u_m)\Big) + 4 \cdot \sum_{l \neq m} \mathrm{tr}(h_{u_l} V^{\otimes n}). \tag{6.32}$$

Combining the definition for the code average error and the bounds in (6.32) together with linearity and monotonicity of the expectation, we arrive at

$$\begin{aligned}
\mathbb{E}_U\Big[\overline{e}(\mathcal{C}(U), V^{\otimes n})\Big] &= \frac{1}{M} \sum_{m=1}^M \mathbb{E}_U\Big[\mathrm{tr} D_m^c(U) V^{\otimes n}(U_m)\Big] \\
&\leq \frac{1}{M} \sum_{m=1}^M \left\{ 2 \cdot \mathbb{E}_{U_m}\Big[\mathrm{tr}(\mathbb{1} - h_{U_m}) V^{\otimes n}(U_m)\Big] + 4 \cdot \sum_{l \neq m} \mathbb{E}_U\Big[\mathrm{tr} h_{U_l} V^{\otimes n}(U_m)\Big] \right\} \\
&= \frac{1}{M} \sum_{m=1}^M \left\{ 2 \cdot \Delta_m + 4 \sum_{l \neq m} \cdot \Lambda_{lm} \right\}.
\end{aligned} \tag{6.33}$$

The last line above is just from setting the abbreviations

$$\Delta_m := \mathbb{E}_{U_m}\Big[\mathrm{tr}(\mathbb{1} - h_{U_m}) V^{\otimes n}(U_m)\Big] \quad \Lambda_{lm} := \mathbb{E}_U\Big[\mathrm{tr} h_{U_l} V^{\otimes n}(U_m)\Big]$$

for each $m \in [M]$, $l \neq m$. We bound the terms separately.

$$\begin{aligned}
\Delta_m &= 1 - \mathbb{E}_{U_m}\left[\mathrm{tr} h_{U_m} V^{\otimes n}(U_m)\right] \\
&= 1 - \sum_{x^n \in \mathcal{X}^n} q^n(x^n)\, \mathrm{tr} h_{x^n} V^{\otimes n}(x^n) \\
&= 1 - \sum_{x^n, y^n \in \mathcal{X}^n} q^n(y^n) \cdot \mathrm{tr}\Big((|e_{x^n}\rangle\langle e_{x^n}| \otimes h_{x^n})(|e_{y^n}\rangle\langle e_{y^n}| \otimes V^{\otimes n}(y^n))\Big) \\
&= 1 - \mathrm{tr}\underbrace{\left(\sum_{x^n \in \mathcal{X}^n} |e_{x^n}\rangle\langle e_{x^n}| \otimes h_{x^n}\right)}_{=\hat{a}}\underbrace{\left(\sum_{y^n \in \mathcal{X}^n} q^n(y^n) |e_{y^n}\rangle\langle e_{y^n}| \otimes V^{\otimes n}(y^n)\right)}_{=\hat{\rho}} \\
&= \mathrm{tr}\hat{a}\hat{\rho} \leq \lambda
\end{aligned} \tag{6.34}$$

The second equality above is by explicitly carrying out the expectation. For the remaining terms we have

$$\begin{aligned}
\Lambda_{lm} &:= \mathbb{E}_U\left[\mathrm{tr} h_{U_l} V^{\otimes n}(U_m)\right] \\
&= E_{U_m U_l}\left[\mathrm{tr} h_{U_l} V^{\otimes n}(U_m)\right] \\
&= \sum_{x^n, y^n \in \mathcal{X}^n} q^n(x^n) q^n(y^n) \cdot \mathrm{tr} h_{x^n} V^{\otimes n}(y^n) \\
&= \sum_{x^n \in \mathcal{X}^n} q^n(x^n) \cdot \mathrm{tr} h_{x^n} \overline{V}_q^{\otimes n} \\
&= \sum_{x^n, z^n \in \mathcal{X}^n} q^n(x^n) \cdot \mathrm{tr}\,(|e_{z^n}\rangle\langle e_{z^n}| \otimes h_{z^n})\left(|e_{x^n}\rangle\langle e_{x^n}| \otimes \overline{V}^{\otimes n}\right) \\
&= \mathrm{tr}\hat{a}\hat{\sigma} \leq \tau.
\end{aligned} \tag{6.35}$$

Using the estimates in (6.34) and (6.35) to further upper bound the right hand side of (6.33), we obtain the inequality in (6.31). As a consequence, we know, that there exists a realization u', such that with the code $\mathcal{C} := \mathcal{C}(u')$ the conclusion $\overline{e}(\mathcal{C}, V^{\otimes n}) < 2\lambda + 4M\tau$ of the proposition is valid. □

Proof of Proposition 79. To prove the coding theorem, we use the construction in Proposition 82 together with Quantum Stein's Lemma (Theorem 30). Fix numbers $\mu \in (0,1), \delta > 0$ and a probability distribution $q \in \mathcal{P}(\mathcal{X})$ sufficient for

$$\chi(q, V) \geq C(V) - \frac{\delta}{4} \tag{6.36}$$

Applying Theorem 30 with hypotheses $\rho := \rho(q, V)$ and $\sigma := \sigma(q, V)$ (the states from (6.28), (6.29)),

it is clear, that there exists a number n_0 such that for all $n > n_0$ we find a test $a_n \in [0, \mathbb{1}^{\otimes n}]$ which fulfills

$$\mathrm{tr} a_n \rho^{\otimes n} \geq 1 - \mu \quad \text{and} \quad \mathrm{tr} a_n \sigma^{\otimes n} \leq \tau_n \tag{6.37}$$

with

$$\tau_n \leq 2^{-n(D(\rho\|\sigma) - \frac{\delta}{2})} = 2^{-n(\chi(q,V) - \frac{\delta}{2})} \leq 2^{-n(C(V) - \frac{3}{4}\delta)} \tag{6.38}$$

The equality above is by the identity from Exercise 81, the rightmost inequality by the choice of q, i.e. (6.36). Applying Proposition 82 with a_n, and M defined by

$$2^{n(C(V)-\frac{\delta}{4})} \geq M := \lfloor 2^{n(C(V)-\frac{\delta}{4})} \rfloor \geq 2^{n(C(V)-\delta)}$$

we know, that there exists an (n, M) code $\mathcal{C}_n$ with

$$\begin{aligned}\overline{e}(\mathcal{C}, V^{\otimes n}) &\leq 2\mu + 4M\tau_n \\ &\leq 2\mu + 4 \cdot 2^{-n\frac{\delta}{2}}.\end{aligned}$$

Setting for given $\lambda \in (0,1)$ we know, from the above reasoning with $\mu = \lambda/2$, that

$$\overline{N}(n, V, \lambda) \geq 2^{n(C(V)-\delta)}$$

for each large enough n. □

6.2 The discrete memoryless quantum channel

In this section, we consider the task of classical message transmission for memoryless quantum channels, which take quantum systems as inputs (and outputs quantum systems.) The noise characteristics of such a channel is completely described by a c.p.t.p. (which explains, why such maps are also called "quantum channels".)

Definition 83. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a c.p.t.p. map. The* discrete memoryless quantum channel (DMQC) *generated by $\mathcal{N}$ is the family*

$$\{\mathcal{N}^{\otimes n} : n \in \mathbb{N}\}. \tag{6.39}$$

The transmission map for n uses of the DMQC $\mathcal{N}$ is $\mathcal{N}^{\otimes n}$.

Next we define message transmission codes for DMQCs.

Definition 84. *An (n, M)-code for transmission of classical messages over the DMQC $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ is a family $\mathcal{C} = (W(m), D_m)_{m=1}^M$ with*

$$W(m) \in \mathcal{S}(\mathcal{H}^{\otimes n}), \quad \text{and}$$

$$D_m \in [0, \mathbb{1}_{\mathcal{K}}^{\otimes n}] \text{ for all } m \in [M] \text{ and } \sum_{m=1}^{M} D_m = \mathbb{1}_{\mathcal{K}}^{\otimes n}.$$

We define the average transmission error *by*

$$\overline{e}(\mathcal{C}, \mathcal{N}^{\otimes n}) := \frac{1}{M} \sum_{m=1}^{M} \mathrm{tr} D_m^c \mathcal{N}^{\otimes n}(W(m)), \tag{6.40}$$

where we again use the notation $D_m^c := \mathbb{1}_{\mathcal{H}}^{\otimes n} - D_m$ for each $m \in [M]$.

As in the case of memoryless classical-quantum channels, we will determine the optimal asymptotical rates for classical message transmission. Therefore, we define for given c.p.t.p. map $\mathcal{N}$ and each $\lambda \in [0,1]$, $n \in \mathbb{N}$

$$\overline{N}(\mathcal{N}, n, \lambda) := \max\left\{M : \exists\ (n, M) - \text{message transmission code } \mathcal{C} \text{ with } \overline{e}(\mathcal{C}, \mathcal{N}^{\otimes n}) \leq \lambda\right\}.$$

Exercise 85 (Maximal error)**.** *In the above definition, we only defined the average transmission error. Define a corresponding maximal error function, and optimal message set sizes $N(\mathcal{N}, n, \lambda)$. Show, that*

$$(1-\lambda^2)\overline{N}(\mathcal{N}, n, \lambda^2) \;\leq\; N(\mathcal{N}, n, \lambda) \;\leq\; \overline{N}(\mathcal{N}, n, \lambda)$$

holds.

Exercise 86. *Show, that for each $\lambda \in (0,1)$*

$$\overline{N}(\mathcal{N}, m, \lambda) \leq \overline{N}(\mathcal{N}, n, \lambda)$$

holds, if $m \leq n$.

Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$. Define for each $k \in \mathbb{N}$,

$$C^{(1)}(\mathcal{N}^{\otimes k}) \;:=\; \sup\{\chi(p, \mathcal{M} \circ V) : \; |\mathcal{Y}| < \infty, \; p \in \mathcal{P}(\mathcal{Y}), \; V : \mathcal{Y} \to \mathcal{S}(\mathcal{H}^{\otimes k})\}. \tag{6.41}$$

We set

$$C(\mathcal{N}) := \sup_{k \in \mathbb{N}} \frac{C^{(k)}(\mathcal{N}^{\otimes k})}{k} \tag{6.42}$$

Next we state the coding theorem and converse for classical message transmission over quantum discrete memoryless channels.

Theorem 87. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$. It holds*

1. $\forall \lambda > 0 : \; \liminf_{n\to\infty} \frac{1}{n} \log \overline{N}(\mathcal{N}, n, \lambda) \;\geq\; C(\mathcal{N})$.
2. $\inf_{\lambda > 0} \limsup_{n\to\infty} \frac{1}{n} \log \overline{N}(\mathcal{N}, n, \lambda) \;\leq\; C(\mathcal{N})$.

Compared to the Holevo information $C(W)$ for a discrete memoryless cq channel defined in (6.3) the above defined capacity function $C(\mathcal{N})$ for a quantum DMC is of more complex structure. While $C(W)$ can be evaluated just by a maximisation problem for the generic cq channel W, $C(\mathcal{N})$ is a so-called *multi-letter formula* which means, that one from principle has to solve a separate optimisation problem for each instance $\mathcal{N}^{\otimes k}$, $k \in \mathbb{N}$. It was for some time a major open question whether or not the problem can be reduced such that $C(\mathcal{N}) = C^{(1)}(\mathcal{N})$ holds, but eventually an example of a channel $\mathcal{N}$ with $C(\mathcal{N}) > C^{(1)}(\mathcal{N})$ was given. However, discussion of the example is beyond the scope of this course. The expression for $C(\mathcal{N})$ can be simplified a bit, i.e. then supremum in (6.42) can be replaced by a limit

Proposition 88. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$. It holds*

$$C(\mathcal{N}) = \lim_{k \in \mathbb{N}} \frac{C^{(k)}(\mathcal{N}^{\otimes k})}{k}$$

To prove the above proposition, we will use the following assertion from real analysis.

Lemma 89 (Fekete's Lemma)**.** *Let $(a_n)_{n=1}^{\infty}$ be a real sequence haveing the property*

$$a_m + a_n \;\leq\; a_{m+n}$$

for all $n, m \in \mathbb{N}$. Then

$$\sup_{n\in\mathbb{N}} \frac{a_k}{k} \leq \lim_{k\to\infty} \frac{a_k}{k}.$$

Proof of Proposition 88. Define

$$a_k := C^{(1)}(\mathcal{N}^{\otimes k}) \qquad\qquad (k \in \mathbb{N}). \tag{6.43}$$

We first convince ourselves, that the sequence $(\frac{1}{k}a_k)_{k=1}^{\infty}$ is convergent. Notice, that the sequence is bounded. It holds

$$0 \leq \frac{a_k}{k} \leq \log\dim\mathcal{K}. \tag{6.44}$$

The left inequality above is obvious from the definitions (the Holevo quantity is always nonnegative. The upper bound above can be seen as follows. For each $k \in \mathbb{N}$, finite alphabet $\mathcal{Y}$, cq channel $V : \mathcal{Y} \to \mathcal{H}^{\otimes k}$ and probability distribution $q \in \mathcal{P}(\mathcal{Y})$, it holds

$$\begin{aligned}
0 \;&\leq \chi(q, \mathcal{N}^{\otimes k} \circ V) \\
&= \; S\left(\overline{(\mathcal{N}^{\otimes k} \circ V)}_q\right) - \sum_{y\in\mathcal{Y}} q(y) S(\mathcal{N}^{\otimes k} \circ V(y)) \\
&\leq \; S\left(\overline{(\mathcal{N}^{\otimes k} \circ V)}_q\right) \\
&\leq \log\dim\mathcal{K}^{\otimes k} \\
&= \; k \log\dim\mathcal{K}.
\end{aligned}$$

The first equality is by definition of the Holevo quantity using the abbreviation

$$\overline{(\mathcal{N}^{\otimes k} \circ V)}_q := \sum_{y\in\mathcal{Y}} q(y) \mathcal{N}^{\otimes k} \circ V(y), \tag{6.45}$$

the last inequality is by the principal bound $S(\rho) \leq \log d$ which holds for each density matrix ρ on a d-dimensional Hilbert space. Maximizing over all finite alphabets and consistent probability distributions and cq channels leads to the desired upper bound.
Next we show subadditivity of the sequence $(\frac{1}{k}a_k)_{k=1}^{\infty}$, i.e.

$$a_m + a_n \leq a_{m+n} \tag{6.46}$$

for all $n, m \in \mathbb{N}$. The above statement together with the bounds in (6.44) show, that the sequence fulfills the conditions of Fekete's Lemma, which implies the desired existence of limit. We show (6.46). Fix any $m, n \in \mathbb{N}$, finite alphabets $\mathcal{Y}$ and $\mathcal{Z}$, and classical-quantum channels $V : \mathcal{Y} \to \mathcal{S}(\mathcal{H}^{\otimes n})$, $W : \mathcal{Z} \to \mathcal{S}(\mathcal{H}^{\otimes n})$. It holds

$$\begin{aligned}
\sup_{p\in\mathcal{P}(\mathcal{Y})} \chi(p, \mathcal{N}^{\otimes m} \circ V) + \sup_{q\in\mathcal{P}(\mathcal{Z})} \chi(p, \mathcal{N}^{\otimes n} \circ W) \;&=\; \sup_{r\in\mathcal{P}(\mathcal{Y}\times\mathcal{Z})} \chi\Big(r, (\mathcal{N}^{\otimes m} \circ V) \otimes (\mathcal{N}^{\otimes n} \circ W)\Big) \\
&=\; \sup_{r\in\mathcal{P}(\mathcal{Y}\times\mathcal{Z})} \chi\Big(r, (\mathcal{N}^{\otimes (m+n)}) \circ (V \otimes W)\Big)
\end{aligned}$$

The first equality above holds by Proposition 76.1. Maximzing both sides of the equality above over all cq channels with finite input alphabets and all compatible probability distributions on the corresponding input alphabets leads to

$$\begin{aligned} a_m + a_n &= C^{(1)}(\mathcal{N}^{\otimes m}) + C^{(1)}(\mathcal{N}^{\otimes n}) \\ &= \sup_{(p,V)} \chi(p, \mathcal{N}^{\otimes m} \circ V) + \sup_{(q,W)} \chi(q, \mathcal{N}^{\otimes m} \circ W) \\ &= \sup_{(p,V)} \sup_{(q,W)} \chi(q, \mathcal{N}^{\otimes m+n} \circ (V \otimes W)) \\ &\leq \sup_{(r,Z)} \chi(r, \mathcal{N}^{\otimes m+n} \circ Z) \\ &= a_{m+n}. \end{aligned}$$

Having convinced ourselves, that the hypotheses of Fekete's Lemma are fulfilled, we can conclude, that

$$\sup_{k \in \mathbb{N}} \frac{C^{(1)}(\mathcal{N}^{\otimes k})}{k}) = \lim_{k \to \infty} \frac{C^{(1)}(\mathcal{N}^{\otimes k})}{k}) \tag{6.47}$$

holds. □

Proof of Theorem 87. We first prove the lower bound in Theorem 87.1. For this reason fix $\lambda > 0$, $k \in \mathbb{N}$, a finite alphabet $\mathcal{Y}$, a probability distribution $p \in \mathcal{P}(\mathcal{Y})$, a classical-quantum channel $V : \mathcal{Y} \to \mathcal{S}(\mathcal{H}^{\otimes k})$. We show, that

$$\liminf_{n \to \infty} \overline{N}(\mathcal{N}, n\lambda) \geq \frac{1}{k} \chi(p, \mathcal{N}^{\otimes k} \circ V)$$

holds. Maximizing over all entities we fixed leads to the desired inequality. Define the effective cq channel $W_{V,k} : \mathcal{Y} \to \mathcal{S}(\mathcal{K}^{\otimes k})$ which arises from concatenating the cq channel V with k instances of $\mathcal{N}$ by

$$W_{V,k}(y) := (\mathcal{N}^{\otimes k} \circ V)(y) = \mathcal{N}^{\otimes k}(V(y)) \qquad (y \in \mathcal{Y}).$$

We first show the inequality

$$\overline{N}(\mathcal{N}, k \cdot l, \lambda) \geq \overline{N}(W_{V,k}, l, \lambda). \tag{6.48}$$

Let $\tilde{\mathcal{C}} := (y_m, D_m)_{m=1}^M$ be an (l, M)-code for $W_{V,k}$. Then $\mathcal{C} := (V^{\otimes l}(y_m), D_m)_{m=1}^M$ is an $(k \cdot l, M)$-code for $\mathcal{N}$. Moreover, it holds for each message $m \in [M]$

$$\mathrm{tr} D_m^c W_{V,k}^{\otimes l}(y_m) = \mathrm{tr}(\mathcal{N}^{\otimes k} \circ V)^{\otimes l}(y_m) D_m^c = \mathrm{tr} D_m^c \mathcal{N}^{\otimes k \cdot l}(V^{\otimes l}(y_m)).$$

By arithmetic averaging over all messages, we arrive at

$$\overline{e}(\tilde{\mathcal{C}}, W_{V,k}^{\otimes l}) = \overline{e}(\mathcal{C}, \mathcal{N}^{\otimes k \cdot l}).$$

We have shown, that we find for each given (l, M) message transmission code for the cq channel $W_{V,k}$ an $(k \cdot l, M)$ message transmission code for the quantum channel $\mathcal{N}$

of the same average error which implies, by maximizing, the inequality in (6.48). Now we can bound

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \overline{N}(\mathcal{N}, n, \lambda) &= \liminf_{l\to\infty} \min_{0\le r<k} \frac{1}{l\cdot k + r} \log \overline{N}(\mathcal{N}, k\cdot l + r, \lambda) \\
&\ge \liminf_{l\to\infty} \min_{0\le r<k} \frac{1}{l\cdot k + r} \log \overline{N}(\mathcal{N}, k\cdot l, \lambda) \\
&= \liminf_{l\to\infty} \frac{1}{k(l+1)} \log \overline{N}(\mathcal{N}, k\cdot l, \lambda) \\
&= \frac{1}{k} \liminf_{l\to\infty} \frac{1}{l+1} \log \overline{N}(\mathcal{N}, k\cdot l, \lambda) \\
&= \frac{1}{k} \liminf_{l\to\infty} \frac{1}{l} \log \overline{N}(\mathcal{N}, k\cdot l, \lambda) \\
&= \frac{1}{k} \liminf_{l\to\infty} \frac{1}{l} \log \overline{N}(W_{V,k}, l, \lambda) \\
&\ge \frac{1}{k} \chi(p, \mathcal{N}^{\otimes k} \circ V).
\end{aligned}$$

The last inequality above is by application of Theorem 73.1 on the cq channel $W_{V,k}$. It remains to show the upper bound in Theorem 87.2. Let $\delta > 0$, and $l_0 \in \mathbb{N}$ an integer large enough such that each $l \ge l_0$ simultaneously fulfills

$$\left| \lim_{k\to\infty} C^{(1)}(\mathcal{N}^{\otimes k}) - \frac{1}{l} C^{(1)}(\mathcal{N}^{\otimes l}) \right| \le \frac{\delta}{2}. \tag{6.49}$$

and $l > \frac{2}{\delta}$. Fix an $l > l_0$ and an (l, M) code for message transmission over $\mathcal{N}$ $\mathcal{C} := (W(m), D_m)_{m=1}^M$ with

$$\overline{e}(\mathcal{C}, \mathcal{N}^{\otimes l}) \le \lambda.$$

Let q_* be the equidistribution on the message set. It holds

$$\begin{aligned}
C^{(1)}(\mathcal{N}^{\otimes l}) &= \sup_{(p,V)} \chi(p, \mathcal{N}^{\otimes l} \circ V) \\
&\ge \chi(q_*, \mathcal{N}^{\otimes l} \circ W) \\
&\ge \vartheta(q_*, \mathcal{N}^{\otimes l} \circ W) \\
&\ge \log M - \lambda \cdot \log M - 1
\end{aligned}$$

Combining the estimates, we have

$$\begin{aligned}
\lim_{k\to\infty} \frac{1}{k} C^{(1)}(\mathcal{N}^{\otimes k}) &\ge \frac{1}{l} C^{(1)}(\mathcal{N}^{\otimes l}) - \frac{\delta}{2} \\
&\ge (1-\lambda)\frac{1}{l} \log M - \frac{1}{l} - \frac{\delta}{2} \\
&\ge (1-\lambda)\frac{1}{l} \log M - \delta.
\end{aligned}$$

Since the code was arbitrary (beside the error threshold), we obtain by maximizing the message set sizes of all codes which fulfill the error threshold

$$\frac{1}{l} \log \overline{N}(\mathcal{N}, l, \lambda) \le \frac{1}{1-\lambda} \left(\lim_{k\to\infty} \frac{1}{k} C^{(1)}(\mathcal{N}^{\otimes k}) + \delta \right)$$

Consequently,

$$\inf_{\lambda>0} \liminf_{l\to\infty} \frac{1}{l} \log \overline{N}(\mathcal{N}, l, \lambda) \;\leq\; \lim_{k\to\infty} \frac{1}{k} C^{(1)}(\mathcal{N}^{\otimes k}) + \delta.$$

Since δ was an arbitrary positive number, we are done. □

Part I

Some more Topics

7 Types and frequency typical sets

Theorem 90. *Let* $\dim \mathcal{K} := d$, $\sigma \in \mathcal{S}(\mathcal{K})$, $\delta \in (0, \frac{1}{2})$. *For each* $k \in \mathbb{N}$ *exists a projector* $p_{k,\delta}(\sigma)$ *such with the following properties.*

1. $p_{k,\delta}(\sigma)\sigma^{\otimes k} = \sigma^{\otimes k} p_{k,\delta}(\sigma)$.
2. $\mathrm{tr}(p_{k,\delta}(\sigma)) \leq \exp\left(k(S(\sigma) + d(\varphi(\delta) + \gamma(k)))\right)$.
3. $\mathrm{tr}(p_{k,\delta}(\sigma)) \geq \exp\left(k(S(\sigma) + d(\varphi(\delta) + \gamma(k)))\right)$.
4. $\mathrm{tr}(p_{k,\delta}(\sigma)\sigma^{\otimes k}) \geq 1 - \exp(-k(c\delta^2 - d\gamma(k)))$.
5. $\exp\left(-k(S(\sigma) + c'(\sigma)\delta)\right) \cdot p_{k,\delta}(\sigma) \leq p_{k,\delta}(\sigma)\sigma^{\otimes k} p_{k,\delta}(\sigma)$.
6. $\exp\left(-k(S(\sigma) - c'(\sigma)\delta)\right) \cdot p_{k,\delta}(\sigma) \geq p_{k,\delta}(\sigma)\sigma^{\otimes k} p_{k,\delta}(\sigma)$,

where functions

$$\varphi : (0, \tfrac{1}{2}] \to \mathbb{R},\ \varphi(t) := -t \log t \qquad (t \in (0, \tfrac{1}{2}) \qquad (7.1)$$

$$\gamma : \mathbb{N} \to \mathbb{R}_+,\ \gamma(n) := \frac{\log(n+1)}{n} \qquad (n \in \mathbb{N}) \qquad (7.2)$$

and the constant $c'(\sigma) := -\mathrm{tr}(\pi \log \sigma)$ *with* π *being the projection onto the support of* σ *where used.*

8 Transmission of classical messages over quantum channels: revisited

In this chapter, we will go somewhat deeper into the issue of ....

Conditionally frequency typical subspaces

In this section, we consider an extension of the notion of frequency typical sets from classical information theory to the setting of cq channels. Let $V : \mathcal{X} \to \infty$ be a cq channel. We abbreviate $d := \dim \mathcal{H}$. To cope with the corresponding notation, we assume that $\mathcal{Y} := \{1, \ldots, d\}$. Let

$$V(x) = \sum_{y \in \mathcal{Y}} \tilde{V}(y|x) |\phi_y(x)\rangle\langle\phi_y(x)| \tag{8.1}$$

be a spectral decomposition of $V(x)$ for each $x \in \mathcal{X}$. To notate the eigenvalue of $V(x)$ belonging to $\phi_y(x)$ by $\tilde{V}(y|x)$ may seem somewhat suggestive, but we have indeed

$$\forall y \in \mathcal{Y} : \ 0 \ \leq \tilde{V}(y|x) \ \leq 1 \quad \text{and} \quad \sum_{y \in \mathcal{Y}} \tilde{V}(y|x) = 1 \tag{8.2}$$

for each $x \in \mathcal{X}$, i.e. the numbers $\{\tilde{V}(y|x)\}_{x \in \mathcal{X}, y \in \mathcal{Y}}$ are the entries of a stocastic $\mathcal{X} \times \mathcal{Y}$ matrix $\tilde{V}$. For each $k \in \mathbb{N}$, and $x^k \in \mathcal{X}^k$,

$$V^{\otimes k}(x^k) = \sum_{y^k \in \mathcal{Y}^k} \tilde{V}^k(y^k|x^k) |\phi_y^k(x^k)\rangle\langle\phi_y^k(x^k)| \tag{8.3}$$

is a spectral decomposition of $V^{\otimes k}(x^k)$, where we used the notation

$$\tilde{V}^k(y^k|x^k) \ = \ \prod_{i=1}^{k} \tilde{V}(y_i|x_i), \text{ and } \quad \phi_{y^k}(x^k) := \phi_{y_1}(x_1) \otimes \cdots \otimes \phi_{y_k}(x_k) \tag{8.4}$$

for each $x^k = (x_1, \ldots, x_k)$, $y^k = (y_1, \ldots, y_k)$. We define for each $\delta > 0$, $x^k \in \mathcal{X}^k$ the *δ-conditional frequency typcial projector to V given x^k by*

$$p_{V,\delta}(x^k) \ := \sum_{y^k \in T_{\tilde{V},\delta}(x^k)} |\phi_{y^k}(x^k)\rangle\langle\phi_{y^k}(x^k)|. \tag{8.5}$$

The following theorem provides some bounds for the frequency typical projections.

Theorem 91. *Let $V : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, $d := \dim \mathcal{H}$, and $\delta < \frac{1}{d|\mathcal{X}|}$, $k \in \mathbb{N}$. It holds*

1. $p_{V,\delta}(x^k) V^{\otimes k}(x^k) \ = \ V^{\otimes k}(x^k) p_{V,\delta}(x^k)$ *for all* $x^k \in \mathcal{X}^k$

2. $\mathrm{tr} p_{V,\delta}(x^k) V^{\otimes k}(x^k) \geq 1 - 2^{-k(c\delta^2 - \alpha_1(k))}$

For each $q \in \mathcal{P}(\mathcal{X})$, $x^k \in T^k_{q,\delta}$, abbreviating $\overline{V}_q := \sum_{x\in\mathcal{X}} q(x)V(x)$, it holds

3. $\mathrm{tr} p_{\overline{V}_q, 2\delta|\mathcal{X}|} V^{\otimes k}(x^k) \geq 1 - 2^{-k(c\delta^2 - \alpha_1(k))}$

4. $2^{k(S(V|q)+\alpha_2(k,\delta))} \geq \mathrm{tr} p_{V,\delta}(x^k) \geq 2^{k(S(V|q)+\alpha_2(k,\delta))}$

5. $2^{-k(S(V|q)-\alpha_3(\delta))} \geq p_{V,\delta}(x^k) V^{\otimes k}(x^k) p_{V,\delta}(x^k) \geq 2^{-k(S(V|q)+\alpha_3(\delta))}$

The strong converse to the coding theorem

In this section, we aim to stregthen the statement from Theorem 73.2. We prove the *strong converse* to the coding theorem 73.1, which is the statement

$$\forall \lambda < 1 : \limsup_{n\to\infty} \frac{1}{n} \log N(n, W, \lambda) \leq \sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p, W). \tag{8.6}$$

Gentle measurement

Lemma 92. *Let $\psi, \phi \in \mathcal{H}$ be unit vectors in $\mathcal{H}$. Then*

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1 = 2\sqrt{1 - |\langle\psi,\phi\rangle|^2}. \tag{8.7}$$

Proof. We may write

$$\phi = \cos(\theta)\psi + e^{i\varphi}\sin(\theta)\psi^\perp \tag{8.8}$$

with some $\theta, \varphi \in [0, 2\pi]$. Notice, that

$$|\langle\psi,\phi\rangle|^2 = \cos^2\theta \tag{8.9}$$

holds. Moreover,

$$|\phi\rangle\langle\phi| = \cos^2\phi\,|\psi\rangle\langle\psi| + \cos\phi\sin\theta e^{i\varphi}\,|\psi\rangle\langle\psi^\perp| + e^{i\varphi}\sin\theta\cos\theta\,|\psi^\perp\rangle\langle\psi| + \sin^2\theta\,|\psi^\perp\rangle\langle\psi^\perp|. \tag{8.10}$$

We now writing $A = |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|$ in matrix form (according to $\psi, \psi^\perp$), we have

$$A = \begin{pmatrix} 1-\cos^2\theta & -e^{-i\varphi}\sin\theta\cos\theta \\ -e^{i\varphi}\sin\theta\cos\theta & -\sin^2\theta \end{pmatrix} = \begin{pmatrix} \sin^2\theta & -e^{-i\varphi}\sin\theta\cos\theta \\ -e^{i\varphi}\sin\theta\cos\theta & -\sin^2\theta \end{pmatrix}.$$

The reader may readily verify, that the eigenvalues of A are $a_1 = |\sin\theta|$ and $a_2 := -|\sin\theta|$. We obtain

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1 = |a_1| + |a_2| = 2\cdot\sqrt{\sin^2\theta} = 2\cdot\sqrt{1-\cos^2\theta} = 2\cdot\sqrt{1-|\langle\psi,\phi\rangle|^2}. \tag{8.11}$$

□

Lemma 93. *Let $a, b \in \mathcal{L}(\mathcal{H})$, $a, b \geq 0$. Then*

$$\|a - b\|_1 \;\geq\; \|a^{\frac{1}{2}} - b^{\frac{1}{2}}\|_2^2. \tag{8.12}$$

Proof. Let

$$a^{\frac{1}{2}} - b^{\frac{1}{2}} = \sum_{i=1}^{d} \mu_i |v_i\rangle\langle v_i| \tag{8.13}$$

be a spectral decomposition of the hermitian matrix $a^{\frac{1}{2}} - b^{\frac{1}{2}}$. It holds

$$\|a^{\frac{1}{2}} - b^{\frac{1}{2}}\|_2^2 \operatorname{tr}(a^{\frac{1}{2}} - b^{\frac{1}{2}})^*(a^{\frac{1}{2}} - b^{\frac{1}{2}}) = \sum_{i=1}^{d} |\mu_i|^2. \tag{8.14}$$

Define a unitary matrix $u \in \mathcal{L}(\mathcal{H})$ by

$$u \;=\; \sum_{i=1}^{d} \operatorname{sign}(\mu_i) |v_i\rangle\langle v_i|, \tag{8.15}$$

where

$$\operatorname{sign}(x) := \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}. \tag{8.16}$$

We have

$$u(a^{\frac{1}{2}} - b^{\frac{1}{2}}) = (a^{\frac{1}{2}} - b^{\frac{1}{2}})u = \sum_{i=1}^{d} |\mu_i| |v_i\rangle\langle v_i| = |a^{\frac{1}{2}} - b^{\frac{1}{2}}|. \tag{8.17}$$

Since for all $y, z \in \mathcal{L}(\mathcal{H})$, the identity

$$x^2 - y^2 = \frac{1}{2}\left((x+y)(x-y) + (x-y)(x+y)\right) \tag{8.18}$$

is valid, we have

$$\begin{aligned}
\|a-b\|_1 &= \max\{|\mathrm{tr}(a-b)w| : \ w \in \mathcal{L}(\mathcal{H}), w^*w = \mathbb{1}_{\mathcal{H}}\} && (8.19)\\
&\geq |\mathrm{tr}(a-b)u| && (8.20)\\
&= |\mathrm{tr}((a^{\frac{1}{2}})^2 - (b^{\frac{1}{2}})^2)u| && (8.21)\\
&= \left|\frac{1}{2}\mathrm{tr}(a^{\frac{1}{2}} + b^{\frac{1}{2}})(a^{\frac{1}{2}} - b^{\frac{1}{2}}) + \frac{1}{2}\mathrm{tr}(a^{\frac{1}{2}} - b^{\frac{1}{2}})(a^{\frac{1}{2}} + b^{\frac{1}{2}})\right| && (8.22)\\
&= \left|\mathrm{tr}(a^{\frac{1}{2}} + b^{\frac{1}{2}})u(a^{\frac{1}{2}} - b^{\frac{1}{2}})\right| && (8.23)\\
&= \mathrm{tr}(a^{\frac{1}{2}} + b^{\frac{1}{2}})|a^{\frac{1}{2}} - b^{\frac{1}{2}}| && (8.24)\\
&= \sum_{i=1}^{d} |\mu_i| \mathrm{tr}(a^{\frac{1}{2}} + b^{\frac{1}{2}})|v_i\rangle\langle v_i| && (8.25)\\
&= \sum_{i=1}^{d} |\mu_i| (\langle v_i, a^{\frac{1}{2}} v_i\rangle + \langle v_i, b^{\frac{1}{2}} v_i\rangle) && (8.26)\\
&\geq \sum_{i=1}^{d} |\mu_i| \cdot |\langle v_i, a^{\frac{1}{2}} v_i\rangle - \langle v_i, b^{\frac{1}{2}} v_i\rangle| && (8.27)\\
&\geq \sum_{i=1}^{d} |\mu_i|^2 && (8.28)\\
&= \|a^{\frac{1}{2}} - b^{\frac{1}{2}}\|_2^2. && (8.29)
\end{aligned}$$

□

Proposition 94. *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be density matrices. It holds*

$$1 - \sqrt{F(\rho,\sigma)} \ \leq \frac{1}{2}\|\rho - \sigma\|_1 \ \leq \ \sqrt{1 - F(\rho,\sigma)}. \tag{8.30}$$

Proof. We first proof the leftmost inequality. It holds

$$\begin{aligned}
\|\rho - \sigma\|_1 &\geq \|\rho^{\frac{1}{2}} - \sigma^{\frac{1}{2}}\|_1 && (8.31)\\
&= \mathrm{tr}(\rho^{\frac{1}{2}} - \sigma^{\frac{1}{2}}) && (8.32)\\
&= \mathrm{tr}\rho - 2\mathrm{tr}\rho^{\frac{1}{2}}\sigma^{\frac{1}{2}} + \mathrm{tr}\sigma && (8.33)\\
&= 2 \cdot (1 - \mathrm{tr}\rho^{\frac{1}{2}}\sigma^{\frac{1}{2}}) && (8.34)\\
&\geq 2 \cdot (1 - \|\rho^{\frac{1}{2}}\sigma^{\frac{1}{2}}\|_1) && (8.35)\\
&= 2 \cdot \left(1 - \sqrt{F(\rho,\sigma)}\right). && (8.36)
\end{aligned}$$

For the remaining inequality, let ψ, ϕ be state vectors of purifications of ρ, σ which fulfill

$$F(\rho,\sigma) = |\langle\phi,\psi\rangle|^2. \tag{8.37}$$

Notice, that such vectors always exist due to Uhlmann's Theorem. We then have

$$\|\rho-\sigma\|_1 \;\leq\; \||\phi\rangle\langle\phi|-|\psi\rangle\langle\psi|\|_1 = 2\sqrt{1-|\langle\phi,\psi\rangle|^2} = 2\sqrt{1-F(\rho,\sigma)}. \tag{8.38}$$

□

Theorem 95. *Let $\rho\in\mathcal{S}(\mathcal{H})$, $E\in\mathcal{L}(\mathcal{H})$, $0\leq E\leq\mathbb{1}_{\mathcal{H}}$. It holds*

$$\|\rho-E^{\frac{1}{2}}\rho E^{\frac{1}{2}}\|_1 \leq 3\sqrt{1-\mathrm{tr}E\rho}. \tag{8.39}$$

Proof. Let $\psi\in\mathcal{H}\otimes\mathcal{H}$ be a state vector of a purification of ρ. Then

$$\psi' := \frac{(E^{\frac{1}{2}}\otimes\mathbb{1}_{\mathcal{H}})|\psi\rangle\langle\psi|(E^{\frac{1}{2}}\otimes\mathbb{1}_{\mathcal{H}})^*}{\sqrt{\langle\psi,(E^{\frac{1}{2}}\otimes\mathbb{1}_{\mathcal{H}})\psi\rangle}} \tag{8.40}$$

is state vector to a purification of

$$\rho' := \frac{E^{\frac{1}{2}}\rho E^{\frac{1}{2}}}{\mathrm{tr}E\rho}. \tag{8.41}$$

Consequently, we have

$$\begin{aligned}
F(|\psi\rangle\langle\psi|,|\psi'\rangle\langle\psi'|) \;&=\; |\langle\psi,\psi'\rangle|^2 && (8.42)\\
&=\; \frac{|\langle\psi,(E^{\frac{1}{2}}\otimes\mathbb{1}_{\mathcal{H}})\psi\rangle|^2}{\langle\psi,(E^{\frac{1}{2}}\otimes\mathbb{1}_{\mathcal{H}})\psi\rangle} && (8.43)\\
&= \mathrm{tr}\,|\psi\rangle\langle\psi|(E^{\frac{1}{2}}\otimes\mathbb{1}_{\mathcal{H}}) && (8.44)\\
&= \mathrm{tr}E^{\frac{1}{2}}\rho && (8.45)\\
&\geq \mathrm{tr}E\rho. && (8.46)
\end{aligned}$$

It follows

$$F(\rho,\rho') \;\geq\; F(|\psi\rangle\langle\psi|,|\psi'\rangle\langle\psi'|) \;\geq\; \mathrm{tr}E\rho \tag{8.47}$$

which implies

$$\|\rho-\rho'\|_1 \;\leq 2\sqrt{1-F(\rho,\rho')} \;\leq\; 2\sqrt{1-\mathrm{tr}E\rho}. \tag{8.48}$$

Finally, we can bound

$$\begin{aligned}
\|\rho-E^{\frac{1}{2}}\rho E^{\frac{1}{2}}\|_1 \;&\leq\; \|\rho-\rho'+\rho'-E^{\frac{1}{2}}\rho E^{\frac{1}{2}}\|_1 && (8.49)\\
&\leq\; \|\rho-\rho'\|_1+\|\rho'-E^{\frac{1}{2}}\rho E^{\frac{1}{2}}\|_1 && (8.50)\\
&\leq\; 2\sqrt{1-\mathrm{tr}E\rho}+|1-\mathrm{tr}E\rho|\cdot\|E^{\frac{1}{2}}\rho E^{\frac{1}{2}}\|_1 && (8.51)\\
&\leq\; 3\sqrt{1-\mathrm{tr}E\rho}. && (8.52)
\end{aligned}$$

□

Definition 96 (Shadow). *Let $\rho \in \mathcal{S}(\mathcal{H})$ be a density matrix, $B \in \mathcal{L}(\mathcal{H})$, $0 \leq B \leq \mathbb{1}_{\mathcal{H}}$, and $\eta \in \mathbb{R} \cap [0,1]$. We call B an η-*shadow *for ρ, if*

$$\mathrm{tr} B\rho \geq \eta. \tag{8.53}$$

Lemma 97 (Shadow Lemma). *Let $E \in \mathcal{L}(\mathcal{H})$, $0 \leq E \leq \mathbb{1}_{\mathcal{H}}$, $\rho \in \mathcal{S}(\mathcal{H})$ such that $\rho E = E\rho$. Assume that*

$$\mathrm{tr} E\rho \geq 1 - \lambda \quad \textit{and} \quad \mu_1 E \leq E^{\frac{1}{2}} \rho E^{\frac{1}{2}} \leq \mu_2 E \tag{8.54}$$

does hold for some positive real numbers λ, μ_1, μ_2. Then

$$\frac{1-\lambda}{\mu_2} \leq \mathrm{tr} E \leq \frac{1}{\mu_1}, \tag{8.55}$$

and, for each $0 \leq B \leq \mathbb{1}_{\mathcal{H}}$ which is an η-shadow for ρ, it holds

$$\mathrm{tr} B \geq \frac{\eta - \lambda}{\mu_2}. \tag{8.56}$$

Proof. Taking traces on both sides of the right hand set of inequalities in (8.54) implies

$$\mu_1 \mathrm{tr} E \leq \mathrm{tr} E\rho \leq \mu_2 \mathrm{tr} E, \tag{8.57}$$

which implies, in turn, using the inequalities $1 - \lambda \leq \mathrm{tr} E\rho \leq 1$, the inequalities in (8.55). To prove (8.56), we bound

$$\begin{aligned}
\mu_2 \mathrm{tr} B \geq \mu_2 \mathrm{tr} BE &\geq \mathrm{tr} BE^{\frac{1}{2}} \rho E^{\frac{1}{2}} && (8.58)\\
&= \mathrm{tr} B\rho - \mathrm{tr}(\rho - E^{\frac{1}{2}} \rho E^{\frac{1}{2}})B && (8.59)\\
&\geq \mathrm{tr} B\rho - \mathrm{tr}(\rho - E^{\frac{1}{2}} \rho E^{\frac{1}{2}}) && (8.60)\\
&\geq \eta - (1 - 1 - \lambda) && (8.61)\\
&= (\eta - \lambda). && (8.62)
\end{aligned}$$

□

Having provided ourselves with the necessary prerequisites, we proceed to prove the strong converse to the coding theorem. We first show, that (subnormalized) codes with all codewords of the same type and small error have suitable upper bounds for their message set sizes.

Proposition 98. *Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, $n \in \mathbb{N}$, $q \in \mathcal{P}(\mathcal{X})$ a type of sequences in $\mathcal{X}^n$. Let $\mathcal{C}_q := (u_m, D_m)_{m=1}^{M_q}$ be a family with*

1. *$u_m \in T_q^n$ for all $m \in [M_q]$*
2. *$0 \leq D_m \leq \mathbb{1}_{\mathcal{H}}^{\otimes n}$ for all $m \in [M_q]$, and $\sum_{m=1}^{M_q} D_m \leq \mathbb{1}_{\mathcal{H}}^{\otimes n}$*
3. *$e_m := \mathrm{tr} D_m^c W^{\otimes n} \leq \lambda$ for all $m \in [M_q]$.*

Then for each $\xi > 0$ there is a number $n_0(\xi, \lambda, W)$ such that for all $n > n_0$ the lower bound

$$M_q \leq \frac{4}{1-\lambda} \exp\left(n(\chi(q, W) + \xi\right) \tag{8.63}$$

is true.

Remark 99. *Notice, that the number n_0 in the above statement is independent of the chosen type q. This will be of some importance when using Proposition 98 to prove a strong converse to Holevo's Theorem*

Proof. Set $d := \dim \mathcal{H}$, and let $\xi > 0$ be fixed. Fix a blocklength $n \in \mathbb{N}$, and a number $\delta > 0$ large enough to simultaneaously fulfill the following three inequalities

$$\exp\left(-n(c\delta^2/2 + \alpha_1(n)/2 - \log 3/n)\right) \leq \frac{1-\lambda}{2}, \tag{8.64}$$

$$\alpha_3(\delta) \leq \xi/2, \tag{8.65}$$

$$\varphi(2|\mathcal{X}|\delta) + \nu(k) \leq \xi/2d. \tag{8.66}$$

We define

$$\overline{W}_q := \sum_{x \in \mathcal{X}} q(x) W(x) \quad \text{and} \quad \hat{q} := p_{n,2|\mathcal{X}|\delta}(\overline{W}_q), \tag{8.67}$$

where $p_{n,2|\mathcal{X}|\delta}(\overline{W}_q)$ is the $2|\mathcal{X}|\delta$-frequency typical projection in $\mathcal{H}^{\otimes n}$ belonging to the average output state $\overline{W}_q$. Moreover, we set

$$W_m := W^{\otimes n}(u_m) \tag{8.68}$$

for each $m \in [M_q]$. We collect some bounds for the above defined Objects, we need later in the proof. By hypothesis, each u_m, $m \in [M_q]$ is of type q, we have

(i) $\operatorname{tr}\hat{q} W_m = \operatorname{tr}\left(p_{n,2|\mathcal{X}|\delta}(\overline{W}_q) W^{\otimes n}(u_m)\right) \geq 1 - 2^{-n(c\delta^2 - \alpha_1(n))}$

(ii) $3\sqrt{1 - \operatorname{tr}\hat{q} W_m} \leq 2^{n(c\delta^2/2 - \alpha_1(n)/2 - \log 3/2)} \leq \frac{1-\lambda}{2}$

where the last inequality above is a consequence of the choice made for n, and δ. Define, for each $m \in [M_q]$ the effect

$$D'_m := \hat{q}^{\frac{1}{2}} D_m \hat{q}^{\frac{1}{2}} \tag{8.69}$$

(remembering that $\hat{q}$ is a projection. It then holds

$$\operatorname{tr} D'_m W_m = \operatorname{tr} D_m \hat{q} W_m \hat{q} \tag{8.70}$$

$$= \operatorname{tr}(D_m W_m) - \operatorname{tr}\left(D_m (W_m - \hat{q} W_m \hat{q})\right). \tag{8.71}$$

The first term on the right hand side of (8.71) is bounded by

$$\operatorname{tr}(D_m W_m) = 1 - \operatorname{tr}(D_m^c W_m) = 1 - e_m \geq 1 - \lambda \tag{8.72}$$

by hypothesis. For the second term, we have

$$\begin{aligned} \mathrm{tr}\,(D_m(W_m - \hat{q}W_m\hat{q})) &\leq \mathrm{tr}(W_m - \hat{q}W_m\hat{q}) && (8.73)\\ &= \|W_m - \hat{q}W_m\hat{q}\|_1 && (8.74)\\ &\leq 3\sqrt{1 - \mathrm{tr}\hat{q}W_m} && (8.75)\\ &\leq \frac{1-\lambda}{2}. && (8.76) \end{aligned}$$

The first inequality and the equality above both are by the fact, that $W_m \geq \hat{q}W_m\hat{q}$ holds. The second inequality is by Lemma 95. The last inequality is by our choice of n and δ. Combining the estimates in (8.76) and (8.72) with the bound in (8.71), we obtain

$$\mathrm{tr}D'_m W_m \ \geq \frac{1-\lambda}{2}. \tag{8.77}$$

Notice, that with $E := p_{W,\delta}(u_m)$, $B := D'_m$ and $\rho := W_m$, the conditions of Lemma 97 are fulfilled, i.e. by Theorem 91, we have

$$\mu_1 E \ \leq \ E^{\frac{1}{2}}\rho E^{\frac{1}{2}} \leq \mu_2 E \tag{8.78}$$

if we set

$$\mu_1 = 2^{-n(S(W|q)+\alpha_3(\delta))} \quad \text{and} \quad \mu_2 = 2^{-n(S(W|q)-\alpha_3(\delta))}. \tag{8.79}$$

and

$$\mathrm{tr}\rho E = \mathrm{tr}W_m D'_m = \mathrm{tr}W^{\otimes n}(u_m)p_{W,\delta}(u_m) \geq 1 - 2^{-n(c\delta^2-\alpha_1(n))}. \tag{8.80}$$

Consequently, we have

$$\mathrm{tr}D'_m \ \geq \left(\frac{1-\lambda}{2} - 2^{-n(c\delta^2-\alpha_1(n))}\right)\cdot \mu_2^{-1} \ \geq \frac{1-\lambda}{4}\cdot\mu_2^{-1}. \tag{8.81}$$

On the other hand,

$$\begin{aligned} \sum_{m=1}^{M_q}\mathrm{tr}D'_m = \ \mathrm{tr}&\left(\hat{q}\sum_{m=1}^{M_q} D_m\right) && (8.82)\\ &\leq \mathrm{tr}\hat{q} && (8.83)\\ &= \mathrm{tr}p_{n,2|\mathcal{X}|\delta}(\overline{W}_q) && (8.84)\\ &\leq 2^{n(S(\overline{W}_q)+d(\varphi(2|\mathcal{X}|\delta)+\nu(n))}. && (8.85) \end{aligned}$$

Consequently,

$$M_q \cdot \frac{1-\lambda}{4}\cdot 2^{n(S(W|q)-\frac{\xi}{2})} \leq \sum_{m=1}^{M_q}\mathrm{tr}D'_m \leq 2^{n(S(\overline{W}_q)+\frac{\xi}{2}} \tag{8.86}$$

rearrangements in the inequality between the leftmost and rightmost terms above yields

$$M_q \leq \frac{1-\lambda}{4}\exp\left(n(S(\overline{W}_q) - S(W|q) + \xi\right) \tag{8.87}$$

which is the claimed upper bound for M_q since $\chi(q,W) = S(\overline{W}_q) - S(W|q)$ by definition of the Holevo quantity. □

We prove the strong converse

Theorem 100. *Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel. It holds for all $\lambda \in (0,1)$*

$$\limsup_{n\to\infty} \frac{1}{n} \log N(n, W, \lambda) \leq \sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p, W). \tag{8.88}$$

Proof. Fix $\lambda, \xi \in (0,1)$, and let $n > n_0$ (where n_0 is the number in the statement of Proposition 98.) Let $\mathcal{C} := (u_m, D_m)_{m=1}^M$ be any (n, M)-code for message transmission over W with maximal error bounded by

$$e(\mathcal{C}, W^{\otimes n}) \leq \lambda. \tag{8.89}$$

Let $\mathcal{T}$ be the set of types in $\mathcal{X}^n$, such that at least one codeword in $\mathcal{C}$ is of that type, i.e.

$$\mathcal{T} := \left\{ q \in \mathcal{T}(n, \mathcal{X}) : \ T_q^n \cap \{u_m\}_{m=1}^M \neq \emptyset \right\}. \tag{8.90}$$

Set

$$\mathcal{M}_q := \{m : \ u_m \in T_q^n\} \tag{8.91}$$

It holds obviously $\mathcal{M}_q \cap \mathcal{M}_{q'} = \emptyset$ if q and q' are not equal. Moreover, $M = \sum_{q\in\mathcal{T}} |\mathcal{M}_q|$. Note, that for each $q \in \mathcal{T}$, the family $(u_m, D_m)_{m\in\mathcal{M}_q}$ fulfills the requirements of Proposition 98. Consequently

$$|\mathcal{M}_q| \ \leq \ \frac{1-\lambda}{4} \cdot 2^{n(\chi(q,W)+\xi)} \ \leq \ \frac{1-\lambda}{4} \cdot 2^{n(\sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p,W)+\xi)} \tag{8.92}$$

Summing up the bound over all $q \in \mathcal{T}$, we have

$$M = \sum_{q\in\mathcal{T}} |\mathcal{M}_q| \tag{8.93}$$

$$\leq |\mathcal{T}| \cdot \frac{1-\lambda}{4} \cdot 2^{n(\sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p,W)+\xi)} \tag{8.94}$$

$$\leq (n+1)^{|\mathcal{X}|} \cdot \frac{1-\lambda}{4} \cdot 2^{n(\sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p,W)+\xi)}. \tag{8.95}$$

Since the code $\mathcal{C}$ was an arbitrary code with maximal error not exceeding λ, we have

$$N(n, W, \lambda) \ \leq \ \frac{1-\lambda}{4} \cdot 2^{n(\sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p,W)+\xi)}, \tag{8.96}$$

and,

$$\limsup_{n\to\infty} \frac{1}{n} \log N(n, W, \lambda) \leq \sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p, W). \tag{8.97}$$

which we set out to prove. □

Alternative proofs for the coding theorem

9 Message identification over classical-quantum channels

In this lecture, we consider identification of classical messages over discrete memoryless classical-quantum channels. The task of message *identification* is to distinguish from that of classical message transmission, which we considered in lecture .. The receiver does not need an answer to the question

What was the message?

Definition 101. *An* $(n, N, \lambda_1, \lambda_2)$ ID code *for the DMCQC* $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ *is a family* $(Q(\cdot|i), D_i)_{i=1}^N$ *such that* $Q(\cdot|i) \in \mathcal{P}(\mathcal{X}^n)$ *is a probability distribution on* $\mathcal{X}^n$ *and* $D_i \in \mathcal{L}(\mathcal{H}^{\otimes n})$, $0 \leq D_i \leq \mathbb{1}$ *for all* $i \in [N]$ *such that*

Definition 102. *An* $(n, N, \lambda_1, \lambda_2)$ ID code *for the DMCQC* $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ *is a family* $(Q(\cdot|i), D_i)_{i=1}^N$ *such that* $Q(\cdot|i) \in \mathcal{P}(\mathcal{X}^n)$ *is a probability distribution on* $\mathcal{X}^n$, $D_i \in \mathcal{L}(\mathcal{H}^{\otimes n})$, $0 \leq D_i \leq \mathbb{1}$ *for all* $i \in [N]$, *and* $\sum_{i=1}^N D_i$ *forms a POVM.*

10 Message transmission over classical-classical-quantum multiple-access channels

In earlier sections it was already considered a situation, where a sender aims to transmit classical messages to a receiver via a memoryless classical-quantum channel. In this section we extend the communication scenario to include a second sender.
We assume presence of a cq channel

$$W : \mathcal{X} \times \mathcal{Y} \ \to \mathcal{S}(\mathcal{H}) \tag{10.1}$$

The *discrete memoryless classical-quantum multiple-access channel* generated by W is the channel model, where the transmission is governed for each blocklength n by the cq channel

$$W^{\otimes n}(x,y) \ := \ \bigotimes_{i=1}^{n} W(y_i, x_i) \tag{10.2}$$

for each $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $y = (y_1, \ldots, y_n) \in \mathcal{Y}^n$.

Example 103. *independent example*

Example 104. *dependent example*

As demonstrated by the second example above, the inputs of on sender can disturb the other senders signal considerably. Below, there is a schematic picture of the general coding scenario in case of a number of n uses of the channel.

Definition 105. *An* (n, M_1, M_2) message transmission code *for the MAC* W *is a family* $\mathcal{C}(u_{1,m_1}, u_{2,m_2} D_{m_1,m_2})_{m_1=1,m_2=1}^{M_1,M_2}$, *where* $u_{1,1}, \ldots, u_{1,M_1} \in \mathcal{X}^n$, $u_{2,1}, \ldots, u_{2,M_2} \in \mathcal{Y}^n$, *and* $(D_{m_1,m_2})_{m_1=1,m_2=1}^{M_1,M_2}$ *is a POVM on* $\mathcal{S}(\mathcal{H})$. *The* (average) transmission error *of* $\mathcal{C}$ *defined by*

$$\overline{e}(\mathcal{C}, W^{\otimes n}) \ := \ \frac{1}{M_1 \cdot M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \mathrm{tr} D_{m_1,m_2}^{c} W^{\otimes n}(m_1, m_2). \tag{10.3}$$

We remember the notation $A^c := \mathbb{1} - A$.

Definition 106. *A pair* (R_1, R_2) *of nonnegative numbers is called an* achievable rate pair *for message transmission over the MAC* W, *if there is a sequence* $(\mathcal{C}_n)_{n \in \mathbb{N}}$, *each* $\mathcal{C}_n$ *being an* $(n, M_{1,n}, M_{2,n})$*-code such that*

1. $\lim_{n \to \infty} \overline{e}(\mathcal{C}_n, W^{\otimes n}) \ = \ 0$,

2. $\liminf_{n\to\infty} \frac{1}{n}\log M_{1,n} \geq R_1$, *and*

3. $\liminf_{n\to\infty} \frac{1}{n}\log M_{2,n} \geq R_2$

We define the message transmission capacity region *of the MAC W by*

$$\overline{C}_{MAC}(W) := \{(R_1, R_2) : (R_1, R_2) \text{ achievable message transmission rate pair}\}$$

The assertions in the following proposition are direct consequences from the definitions made.

Proposition 107. *For each cq MAC W, the capacity region $\overline{C}_{MAC}(W)$ is a compact and convex subset of $\mathbb{R}^2$.*

Theorem 108. *Let $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H})$ be a cq channel. It holds*

$$\overline{C}_{MAC}(W) = \bigcup_{\substack{p\in\mathcal{P}(\mathcal{X})\\ q\in\mathcal{P}(\mathcal{Y})}} \mathcal{R}(p,q), \tag{10.4}$$

we will prove Theorem 109 in two portions. Proposition ... formulates the coding theorem, while the converse is Proposition

Proposition 109. *Let $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H})$ be a cq channel. It holds*

$$\overline{C}_{MAC}(W) \supset \bigcup_{\substack{p\in\mathcal{P}(\mathcal{X})\\ q\in\mathcal{P}(\mathcal{Y})}} \mathcal{R}(p,q), \tag{10.5}$$

Proposition 110. *Let $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H})$ be a cq channel. It holds*

$$\overline{C}_{MAC}(W) \subset \bigcup_{\substack{p\in\mathcal{P}(\mathcal{X})\\ q\in\mathcal{P}(\mathcal{Y})}} \mathcal{R}(p,q), \tag{10.6}$$

Proof. Let $p \in \mathcal{P}(\mathcal{X}), q \in \mathcal{P}(\mathcal{Y})$ be arbitrary and fixed probability distributions. We show the inclusion $\overline{C}_{MAC}(W) \subset \mathcal{R}(p,q)$.
Fix an arbitrary number $\delta > 0$. We show achievability for the rate pair (R_1, R_2),

$$R_1 := I(A;C) - \delta \tag{10.7}$$
$$R_2 := I(B;C|A) - \delta \tag{10.8}$$

Fix $n \in \mathbb{N}$. We will randomly choose codewords for each sender and use the random coding lemma ... . We introduce two independent families of i.i.d. random variables

$$U := (U_1, \ldots, U_{M_1}) \text{ with } \Pr(U_i = x^n) = p^n(x^n) \text{ for each } x^n \in \mathcal{X}^n \tag{10.9}$$
$$V := (V_1, \ldots, V_{M_2}) \text{ with } \Pr(V_i = y^n) = q^n(y^n) \text{ for each } y^n \in \mathcal{Y}^n. \tag{10.10}$$
$$\tag{10.11}$$

We define for each realization v of V a cq channel

$$W_{1,v} : \mathcal{X}^n \to \mathcal{S}(\mathcal{H}^{\otimes n}),\ x \mapsto W_{1,y}(x^n) := \frac{1}{M_2} \sum_{m_2=1}^{M_2} W^{\otimes n}(x^n, v_{m_2}) \tag{10.12}$$

and for each realization v of V a cq channel

$$W_{2,u} : \mathcal{Y}^n \to \mathcal{S}(\mathcal{H}^{\otimes n}),\ y \mapsto W_{2,x}(y^n) := \frac{1}{M_1} \sum_{m_1=1}^{M_1} |u_{m_1}\rangle\langle u_{m_1}| \otimes W^{\otimes n}(u_{m_1}, y^n) \tag{10.13}$$

□

11 A Chernov-Hoeffding type Concentration inequality for random matrices

In this section, we prove a matrix generalization of the well-known classical Chernov bound, which reads as follows.

Theorem 111 (Lieb's Theorem)**.** *Let $H \in \mathbb{H}_d$. The matrix trace function*

$$A \mapsto \operatorname{tr} \exp(H + \log A) \tag{11.1}$$

is concave on $\mathbb{H}_d^{++}$

Definition 112 (Matrix relative entropy)**.** *Let $A, B \in \mathbb{H}_d^{++}$. The matrix relative entropy of (A, B) is defined by*

$$D(A\|B) \;:=\; \operatorname{tr}\left(A(\log A - \log B) - (A - B)\right). \tag{11.2}$$

Note that the matrix relative entropy defined above is nothing else, than the extension of the quantum relative entropy defined in Def. 29 to the set of all nonnegative matrices (which are invertible). In fact, let $A, B \in \mathbb{H}_d^{++}$, be nonnegative matrices with

$$A \;=\; \alpha \cdot \rho \quad \text{and} \quad B \;=\; \beta \cdot \sigma \tag{11.3}$$

and $\rho, \sigma \in \mathcal{S}(\mathbb{C}^d)$ full-rank density matrices, then

$$\begin{aligned} D(A\|B) \;&=\; D(\alpha\rho\|\alpha\sigma) && (11.4)\\ &=\; \operatorname{tr}(\alpha\rho(\log\alpha\rho) - \log(\beta\sigma)) - \operatorname{tr}(\alpha\rho - \beta\sigma) && (11.5)\\ &=\; \operatorname{tr}(\alpha\rho(\log\alpha\mathbb{1}) + \log\rho) - \log(\beta\mathbb{1})) - \log(\sigma)) - (\alpha - \beta) && (11.6)\\ &=\; \alpha D(\rho\|\sigma) + \alpha\log\frac{\alpha}{\beta} - (\alpha - \beta). && (11.7) \end{aligned}$$

In consequence, some important properties of the quantum relative entropy an be concluded to the matrix relative entropy. An example of this fact is the monotonicity under completely positive and trace preserving maps, which will be the starting point for the rest of the chapter, and was already shown for density matrix arguments in Theorem 74. A restatement for the present case reads as follows.

Corollary 113 (CPTP Monotonicity of the Matrix relative entropy)**.** *Let $A, B \in \mathbb{H}_n^+$+, $\mathcal{N}; \mathcal{L}(\mathbb{C}^n) \to \mathcal{L}(\mathbb{C}^m)$ be a c.p.t.p. map. It holds*

$$D(\mathcal{N}(A)\|\mathcal{N}(B)) \leq D(A\|B) \tag{11.8}$$

Proof. We have

$$D(A\|B) - D(\mathcal{N}(A)\|\mathcal{N}(B)) = \alpha(D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))) \geq 0. \tag{11.9}$$

The equality above is by Eq. (11.7). The inequality follows by positivity of α and Theorem 74. □

Theorem 114 (Joint Convexity of the matrix relative entropy)**.** *Let* $A_1, A_2, B_1, B_2 \in \mathbb{H}_d^{++}$*, and* $\lambda \in (0,1)$*. It holds*

$$D(\lambda A_1 + (1-\lambda)A_2 \| \lambda B_1 + (1-\lambda)B_2) \leq \lambda D(A_1\|B_1) + (1-\lambda)D(A_2\|B_2) \tag{11.10}$$

Proof. Define nonnegative matrices

$$A := \lambda A_1 \otimes |e_1\rangle\langle e_1| + (1-\lambda)A_2 \otimes |e_2\rangle\langle e_2| \quad \text{and} \tag{11.11}$$

$$B := \lambda B_1 \otimes |e_1\rangle\langle e_1| + (1-\lambda)B_2 \otimes |e_2\rangle\langle e_2|. \tag{11.12}$$

We have

$$D(A\|B) = \lambda D(A_1\|B_1) + (1-\lambda)D(A_2\|B_2), \tag{11.13}$$

on the other hand, tracing out the additional systems gives

$$\mathrm{tr}_{\mathbb{C}^2} A = \lambda A_1 + (1-\lambda)A_2, \text{ and} \mathrm{tr}_{\mathbb{C}^2} B = \lambda B_1 + (1-\lambda)B_2. \tag{11.14}$$

In combination, we have

$$\begin{aligned} \lambda D(A_1\|B_1) + (1-\lambda)D(A_2\|B_2) &= D(A\|B) && (11.15)\\ &\geq D(\mathrm{tr}_{\mathbb{C}^2} A \| \mathrm{tr}_{\mathbb{C}^2} B) && (11.16)\\ &= D(\lambda A_1 + (1-\lambda)A_2 \| \lambda B_1 + (1-\lambda)B_2). && (11.17) \end{aligned}$$

The inequality above is by c.p.t.p. monotonicity of the matrix relative entropy. □

Lemma 115. *Let* $f : \mathbb{H}_d^{++} \times \mathbb{H}_d^{++} \to \mathbb{R}$ *be jointly concave. The function* $\tilde{f} : \mathbb{H}_d^{++} \to \mathbb{R}$*, given by*

$$\tilde{f}(B) = \sup_{A \in \mathbb{H}_d^{++}} f(A,B) \qquad (B \in \mathbb{H}_d^{++}) \tag{11.18}$$

is concave.

Proof. Fix $\epsilon > 0$. For each $B_1, B_2 \in \mathbb{H}_d^{++}$ exist $A_1, A_2 \in \mathbb{H}_d^{++}$ such that

$$f(A_1, B_1) \geq \sup_{A \in \mathbb{H}_d^{++}} f(A, B_1) - \epsilon = \tilde{f}(B_1) - \epsilon, \text{ and} \tag{11.19}$$

$$f(A_2, B_2) \geq \sup_{A \in \mathbb{H}_d^{++}} f(A, B_2) - \epsilon = \tilde{f}(B_2) - \epsilon \tag{11.20}$$

For each $\lambda \in (0,1)$, we have

$$\begin{aligned} \tilde{f}(\lambda B_1 + (1-\lambda)B_2) &= \sup_{A \in \mathbb{H}_d^{++}} f(A, \lambda B_1 + (1-\lambda)B_2) && (11.21)\\ &\geq f(\lambda A_1 + (1-\lambda)A_2, \lambda B_1 + (1-\lambda)B_2) && (11.22)\\ &\geq \lambda f(A_1, B_1) + (1-\lambda) f(A_2, B_2) && (11.23)\\ &\geq \lambda(\tilde{f}(A_1) - \epsilon) + (1-\lambda)(\tilde{f}(B_2) - \epsilon). && (11.24) \end{aligned}$$

Since ϵ was an arbitrary positive number, we are done. □

Lemma 116. *Let $\mathbb{M} \in \mathbb{H}_d^{++}$. It holds*

$$\mathrm{tr}M = \sup_{T \in \mathbb{H}_d^{++}} \mathrm{tr}(T \log M - T \log T + T) \tag{11.25}$$

Proof. By rearrangement of the inequality $D(M||T) \geq 0$, we know, that the l.h.s. of the equality to prove exceeds the r.h.s. On the other hand, equality is attained when $T = M$. □

Proof of Theorem 111. We use Lemma 116 with $M := \exp(H + \log A)$ to obtain

$$\begin{aligned} \mathrm{tr}\exp(H + \log A) &= \sup_{T \in \mathbb{H}_d^{++}} \mathrm{tr}\,(T(H + \log A) - T \log T + T) && (11.26)\\ &= \sup_{T \in \mathbb{H}_d^{++}} \mathrm{tr}\,(TH) + \mathrm{tr}(A) - D(T||A)). && (11.27)\\ & && (11.28) \end{aligned}$$

Since the matrix relative entropy is jointly convex, the function

$$f(T, A) := \mathrm{tr}(TH) + \mathrm{tr}A - D(T||A) \tag{11.29}$$

is jointly concave. Using Lemma 115, and the variational formula above, we are done. □

Next, we aim to prove a chernov-type bound for sums of independent hermitian random matrices. For each hermitian matrix $A \in \mathbb{H}_d$, we use the shorcuts $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote the maximal and minimal eigenvalues. It holds

Proposition 117. *Let $Y : \Omega \to \mathbb{H}_d$ be a hermitian random matrix. For all $t \in \mathbb{R}$, it holds*

$$\begin{aligned} \Pr(\lambda_{\max}(Y) \geq t) &\leq \inf_{\theta > 0} e^{-\theta t} \cdot \mathbb{E}\mathrm{tr}e^{\theta Y}, \quad \textit{and} && (11.30)\\ \Pr(\lambda_{\min}(Y) \leq t) &\leq \inf_{\theta < 0} e^{-\theta t} \cdot \mathbb{E}\mathrm{tr}e^{\theta Y}. && (11.31) \end{aligned}$$

Theorem 118. *Let $\{X_k\}_{k=1}^N$ be an independent family of hermitian random matrices, $X_k : \Omega \to \mathbb{H}_d$ ($k \in [N]$). Set $Y := \sum_{k=1}^N X_k$. It holds for all $t \in \mathbb{R}$*

12 Transmission of private messages over quantum channels

13 The genuine quantum capacities

14 Teleportation and Dense Coding

In this chapter, we consider *teleportation* and *dense coding* two prominent protocols being prominent in quantum information theory.

Local operations and classical communication

We introduce notions which go beyond the concept of quantum channels.

Definition 119 (Operation)**.** *Let $\mathcal{H},\mathcal{K}$ be Hilbert spaces. A completely positive map $\mathcal{T}:\ \mathcal{L}(\mathcal{H})\to\mathcal{L}(\mathcal{K})$ which in addition is trace-nonincreasing, i.e.*

$$\mathrm{tr}\mathcal{T}(A)\leq \mathrm{tr}A \qquad\qquad (A\in\mathcal{L}(\mathcal{H}), A\geq 0) \tag{14.1}$$

is called (quantum) operation*. We define the shortcut*

$$\mathcal{C}^{\downarrow}(\mathcal{H},\mathcal{K}) := \{\mathcal{T}:\mathcal{L}(\mathcal{H})\to\mathcal{L}(\mathcal{K}):\ \mathcal{T}\ \textit{is c.p., and}\ \forall A\geq 0:\ \mathrm{tr}\mathcal{T}(A)\leq \mathrm{tr}A\}. \tag{14.2}$$

Remark 120. *(i) By definition, a quantum channel (completely positive and trace preserving map) is an operation, i.e. $\mathcal{C}(\mathcal{H},\mathcal{K})\subset\ \mathcal{C}^{\downarrow}(\mathcal{H},\mathcal{K})$.*

(ii) Because an operation is c.p. in particular, it admits a Kraus representation.

(iii) An operation can be always completed to be a quantum channel by adding a suitable c.p. map.

Definition 121 (Instrument)**.** *Let $|\mathcal{X}|<\infty$. A* (quantum) instrument *is a family $\{\mathcal{T}_x\}_{x\in\mathcal{X}}$ such that*

1. *$\mathcal{T}_x\in\mathcal{C}^{\downarrow}(\mathcal{H},\mathcal{K})$ for each $x\in\mathcal{X}$, and*
2. *$\sum_{x\in\mathcal{X}}\mathcal{T}_x\in\mathcal{C}(\mathcal{H},\mathcal{K})$.*

Definition 122 (One-way LOCC channels)**.** *Let $\mathcal{H}_A,\mathcal{H}_B,\mathcal{K}_A,\mathcal{H}_B$ be Hilbert spaces of systems under control of communication parties A and B. A quantum channel $\mathcal{N}\in\mathcal{C}(\mathcal{H}_A\otimes\mathcal{H}_B,\mathcal{K}_A\otimes\mathcal{K}_B)$ is a LOCC channel with local operations regarding A and B and (noiseless) classical communication from A to B ($A\to B$-one-way LOCC channel), if it can be written in the form*

$$\mathcal{N}(a) := \sum_{k=1}^{N}\mathcal{A}_k\otimes\mathcal{B}_k(a) \tag{14.3}$$

where $\{\mathcal{A}_k : k\in[N]\}\subset\mathcal{C}\downarrow(\mathcal{H}_A,\mathcal{K}_A)$ is a quantum instrument, and $\mathcal{B}_k\in\mathcal{C}(\mathcal{H}_B,\mathcal{K}_B)$ is a quantum channel for each $k\in[N]$.

Entanglement-enhanced LOCC - Quantum Teleportation

It is known, that one cannot exceed the class of separable states by LOCC mappings. A very interesting class of protocols arises, if two parties can use preshared entangled states in addition to local operations and classical communiation. One prominent example of this class is the so-called *quantum teleportation* protocol.

Theorem 123 (Quantum teleportation)**.** *Let $\mathcal{H}_A \simeq \mathcal{K}_A \simeq \mathcal{K}_B$ be Hilbert spaces. There exists an $A \to B$ one-way LOCC channel $\mathcal{T} \in \mathcal{C}(\mathcal{H}_A \otimes \mathcal{K}_A \otimes \mathcal{K}_B, \mathcal{K}_B)$ such that with a pure maximally entangled state $\Phi \in \mathcal{S}(\mathcal{K}_A \otimes \mathcal{K}_B)$*

$$\mathcal{T}(a \otimes \Phi) \;=\; a \tag{14.4}$$

for each $a \in \mathcal{L}(\mathcal{H}_A)$.

Before we prove the above assertion, we supply us with the following lemma.

Lemma 124. *Let $d \in \mathbb{N}$. There exists a family $\{v_i\}_{i=1}^{d^2} \subset \mathcal{L}(\mathbb{C}^d)$ of unitary matrices, such that the following properties hold.*

1. $\mathrm{tr} v_i^* v_k \;=\; d\delta_{ij}$ *for all* $i, j \in [d^2]$
2. *With* $\phi := \sqrt{d}^{-1} \sum_{k=1}^{d} e_k \otimes e_k$ *and* $\phi_i := (v_i \otimes \mathbb{1})\phi$ *for each* $i \in [d]$*, it holds*
   - 2a. $\langle \phi_i, \phi_j \rangle \;=\; \delta_{ij}$ *for all* $i, j \in [d]$
   - 2b. $\sum_{i=1}^{d^2} |\phi_i\rangle\langle\phi_i| \;=\; \mathbb{1} \otimes \mathbb{1}$.

Proof. Let $\gamma : \{0, d-1\} \times \{0, d-1\} \to [d^1]$ be any bijection. We will show, that the matrices

$$v_i := v_{\gamma(r,s)} := \sum_{l=0}^{d-1} e^{-i2\pi sl/d} |e_{l \ominus r}\rangle\langle e_l| \qquad (i \in [d^2]) \tag{14.5}$$

suffice the properties claimed in the lemma ($\ominus$ denotes the modulo-d substraction). Unitarity of $v_1, \ldots, v_{d^2}$ follows by straightforward calculation. To show the first claim of the lemma, let $(r, s) = \gamma - 1(j)$, $(r', s') = \gamma^{-1}(k)$. It holds

$$\mathrm{tr} v_k^* v_j \;=\; \sum_{l,l'=0}^{d-1} e^{i2\pi'(s'l' - sl)/d} \delta_{r'r} \cdot \delta_{l'l} \;=\; \sum_{l=0}^{d-1} e^{i2\pi(s'-s)l/d} \delta_{rr'} \tag{14.6}$$

We assume $r = r'$ and evaluate the right hand side of (16.1). If $s = s'$, one can directly see, that $j = k$, and

$$\mathrm{tr}_{v_k^* v_j} = d. \tag{14.7}$$

When $s \neq s'$, then

$$\sum_{l=0}^{d-1} e^{i2\pi(s'-s)l/d} \;=\; \sum_{l=0}^{d-1} (e^{i2\pi(s'-s)/d})^l = 0 \tag{14.8}$$

To verify the rightmost of the above inequalites, we notice, that evaluating the geometric sum $\sum_l = 0^{d-1} q^l$ with $q = e^{i2\pi(s'-s)/d}$ leads us to

$$\sum_{l=0}^{d-1} (e^{i2\pi(s'-s)/d})^l = \frac{1 - e^{i2\pi(s-s')}}{1 - e^{(s-s')/d}} = 0. \tag{14.9}$$

We summarize

- $k \neq j \;\Rightarrow\; r \neq r' \vee s \neq s' \;\Rightarrow \mathrm{tr} v_k^* v_j = 0$
- $k = j \;\Rightarrow\; r = r' \wedge s = s' \;\Rightarrow \mathrm{tr} v_k^* v_j = d$

which shows the first claim. The remaining statements are rather straightforward applications of the first statement and are left as exercises. □

Proof of Theorem 123. Let $\dim \mathcal{H}_A = d$, and $\Phi := |\phi\rangle\langle\phi|$ with $\phi := \sqrt{d}^{-1} \sum_{k=1}^{d} e_k \otimes e_k$. Using the family $\{v_i\}_{i=1}^{d^2}$ and the family $\{\phi_i\}_{i=1}^{d^2}$ from lemma 124 respectively, we set $\Phi_k := |\phi_k\rangle\langle\phi_k|$,

$$\mathcal{E}_k(\cdot) := P_k(\cdot)P_k^* \in \mathcal{C}^{\downarrow}(\mathcal{H}_A \otimes \mathcal{K}_A, \mathcal{H}_A \otimes \mathcal{K}_A) \quad \text{and} \tag{14.10}$$

$$\mathcal{U}_k(\cdot) := v_k(\cdot)v_k^* \in \mathcal{C}(\mathcal{K}_B, \mathcal{K}_B) \tag{14.11}$$

for each $k \in [d^2]$. It is easy to check, usign the claims of Lemma 124, that $\{\mathcal{E}_k : k \in [d^2]\}$ is a quantum instrument, and consequently

$$\tilde{\mathcal{T}}(\cdot) := \sum_{k=1}^{d^2} \mathcal{E}_k \otimes \mathcal{U}_k(\cdot) \tag{14.12}$$

is an $A \to B$ one-way LOCC channel. Moreover, the channel $\mathcal{T} := \mathrm{tr}_{\mathcal{H}_A \otimes \mathcal{K}_A} \circ \tilde{\mathcal{T}}$ is also an $A \to B$ one-way LOCC channel (Check this!). We show, that $\mathcal{T}$ suffices the identity in (14.4). Since the map $a \mapsto \mathcal{T}(a \otimes \Phi)$ is linear, it is enough to show the identity for all matrix units $E_{ij} := |e_i\rangle\langle e_j|$, $i, j \in [d]$. With the notation introduced, notice, that

$$\Phi = \frac{1}{d} \sum_{i,j=1}^{d} E_{ij} \otimes E_{ij}, \text{ and} \tag{14.13}$$

$$\Phi_k = \frac{1}{d} \sum_{i,j=1}^{d} E_{ij} \otimes v_k E_{ij} v_k^* \qquad (k \in [d^2]) \tag{14.14}$$

holds. We individually evaluate the summands in the right hand side of Eq. (14.12). It holds for each $k \in [d^2], m, n \in [d]$

$$\mathcal{E}_k \otimes \mathcal{U}_k(E_{mn} \otimes \Phi) = (\Phi_k \otimes v_k^*)(E_{mn} \otimes \Phi)(\Phi_k \otimes v_k^*)^* \tag{14.15}$$

$$= \frac{1}{d} \sum_{i,j=1}^{d} (\Phi_k \otimes v_k^*)(E_{mn} \otimes E_{ij} \otimes E_{ij})(\Phi_k \otimes v_k^*)^* \tag{14.16}$$

$$= \Phi_k \otimes \left(\frac{1}{d} \sum_{i,j=1}^{d} \langle \phi_k, e_m \otimes e_i \rangle \langle e_n \otimes e_k, \phi_k \rangle \cdot v_k^* E_{ij} v_k \right) \tag{14.17}$$

A calculation shows, that the termin in brackets above equals $\frac{1}{d}E_{mn}$. Summing up the equalities, we have

$$\begin{aligned}
\mathcal{T}(\Phi, E_{mn}) &= \mathrm{tr}_{\mathcal{H}_A \otimes \mathcal{K}_A}\left(\sum_{k=1}^{d^2} \mathcal{E}_k \otimes \mathcal{U}_k(\Phi \otimes E_{mn})\right) && (14.18)\\
&= \mathrm{tr}_{\mathcal{H}_A \otimes \mathcal{K}_A}\left(\sum_{k=1}^{d^2} \Phi_k \otimes E_{mn}\right) && (14.19)\\
&= E_{mn}. && (14.20)
\end{aligned}$$

□

15 Entanglement Cost

16 Entanglement Assisted Classical Message Transmission

In the preceding lectures, we have discussed the task of classical message transmission over discrete memoryless quantum channels. We noticed, that we end up with a so-called multi-letter formula for the corresponding capacity which turns out as very unsatisfactory when it comes to calculating capacities.
A very interesting scenario arises, if sender and receiver have access to shared entanglement in addition to the quantum channel. In case, that enough entanglement is present, the corresponding capacity turns out to be described by a single-letter formula, as we will see in this chapter. After all, the capacity formula we derive shows, that sometimes entanglement helps to achieve higher classical message transmission capacities for channel transmission. This may be regarded as somewhat surprising, since shared entanglement is not sufficient as resource for any nontrivial transmission of classical messages.
Before we introduce precise definitions for the above mentioned coding scenario, we provide ourselves with a formalized evidence for the claim "shared entanglement alone does not suffice for message transmission".
Let A denote the sending party, while B is the receiver. Assume, the share a state $\rho \in \mathcal{S}(\mathcal{K}_A \otimes \mathcal{K}_B)$. The most general way to set up a message transmission scheme for a number of M messages is to assign a c.p.t.p. map $\mathcal{E}_m : \mathcal{L}(\mathcal{K}_A) \to \mathcal{L}(\mathcal{H}_A)$ with any Hilbert space $\mathcal{H}_{A'}$, and a matrix D_m, $0 \leq D_m \leq \mathbb{1}_{\mathcal{H}_B}$ for each $m \in [M]$, such that $\sum_{m=1}^{M} D_m = \mathbb{1}_{\mathcal{H}_B}$. Assume, that

$$\mathcal{E}_m(a) = \sum_{k=1}^{K} A_k a A_k^* \qquad\qquad (a \in \mathcal{L}(\mathcal{K}_A))$$

is any Kraus decomposion for $\mathcal{E}_m$. The probability, that m' is received, while m was sent, is given by

$$\begin{aligned} p(m'|m) &= \operatorname{tr}(\mathbb{1}_{\mathcal{H}_A} \otimes D_{m'})(\mathcal{E}_m \otimes \mathrm{id}_{\mathcal{H}_B}(\rho)) \\ &= \sum_{k=1}^{K} \operatorname{tr}(\mathbb{1}_{\mathcal{H}_A} \otimes D_{m'})(A_k \otimes \mathbb{1}_{\mathcal{H}_B})(\rho)(A_k \otimes \mathbb{1}_{\mathcal{H}_B})^* \\ &= \operatorname{tr}\left(\sum_{k=1}^{K} A_k^* A_k\right) \otimes D_{m'}\rho \\ &= \operatorname{tr} D_{m'}\rho_B. \end{aligned}$$

Inspection of the above chain of equalities shows, that the probability to receive m' is *independent of the sent* m. In consequence, if $m_1, m_2 \in [M]$ are any two distinct messages, and $p(m_1|m_1) \geq 1 - \lambda$ for some $\lambda \in [0, 1]$, then

$$p(m_2|m_2) \;=\; 1 - \sum_{m \neq m_2} p(m|m_2) \;\leq\; 1 - p(m_1|m_2) \;=\; 1 - p(m_1|m_1) \;<\; \lambda.$$

It is therefore not possible, to transmit one of two messages with an error less than $\frac{1}{2}$, which can be also achieved if the receiver randomly guesses the message.
Another "extreme" case arises from using the ideal quantum channel $\mathrm{id}_{\mathcal{H}}$ together with a pure maximally entangled state on $\mathcal{H}\otimes\mathcal{H}$. In this case, using the so-called *dense coding protocol*, d^2 messages can be transmitted with perfect reliability, i.e. the message transmission capacity without entanglement assistance is exceeded by a factor of 2!

Theorem 125 (Dense coding)**.** *Let $\mathcal{H}_A = \mathcal{H}_B = \mathcal{K}_B = \mathbb{C}^d$,* id *the ideal channel mapping $\mathcal{H}_A$ to $\mathcal{H}_B$. There exists a family $\{\hat{\mathcal{E}}_m\}_{m=1}^{d^2} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_A)$, and a POVM $\{\hat{D}_m\}_{m=1}^{d^2}$ such that with $\phi := \sqrt{d}^{-1}\sum_{k=1}^{d} e_k \otimes e_k \in \mathcal{H}_A \otimes \mathcal{K}_B$ for each $m, m' \in [d^2]$*

$$p(m'|m) \;:=\; \mathrm{tr}\Big(\mathrm{id}\circ\hat{\mathcal{E}}_m \otimes \hat{D}_{m'}|\phi\rangle\langle\phi|\Big) \;=\; \delta_{mm'}.$$

To prove the above assertion, we will use the following proposition

Lemma 126. *Let $d \in \mathbb{N}$. There exists a family $\{v_i\}_{i=1}^{d^2} \subset \mathcal{L}(\mathbb{C}^d)$ of unitary matrices, such that the following properties hold.*

1. $\mathrm{tr} v_i^* v_k \;=\; d\delta_{ij}$ *for all $i, j \in [d^2]$*
2. *With $\phi := \sqrt{d}^{-1}\sum_{k=1}^{d} e_k \otimes e_k$ and $\phi_i := (v_i \otimes \mathbb{1})\phi$ for each $i \in [d]$, it holds*
   - 2a. $\langle\phi_i, \phi_j\rangle \;=\; \delta_{ij}$ *for all $i, j \in [d]$*
   - 2b. $\sum_{i=1}^{d^2} |\phi_i\rangle\langle\phi_i| \;=\; \mathbb{1}\otimes\mathbb{1}$.

Proof. Let $\gamma : \{0, d-1\}\times\{0, d-1\} \to [d^2]$ be any bijection. We will show, that the matrices

$$v_i := v_{\gamma(r,s)} := \sum_{l=0}^{d-1} e^{-i2\pi sl/d}\,|e_{l\ominus r}\rangle\langle e_l| \qquad\qquad (i \in [d^2])$$

suffice the properties claimed in the lemma ($\ominus$ denotes the modulo-d substraction). Unitarity of $v_1, \ldots, v_{d^2}$ follows by straightforward calculation. To show the first claim of the lemma, let $(r, s) = \gamma - 1(j)$, $(r', s') = \gamma^{-1}(k)$. It holds

$$\mathrm{tr} v_k^* v_j \;=\; \sum_{l,l'=0}^{d-1} e^{i2\pi'(s'l'-sl)/d}\delta_{r'r}\cdot\delta_{l'l} \;=\; \sum_{l=0}^{d-1} e^{i2\pi(s'-s)l/d}\delta_{rr'}$$

We assume $r = r'$ and evaluate the right hand side of (16.1). If $s = s'$, one can directly see, that $j = k$, and

$$\mathrm{tr}_{v_k^* v_j} = d.$$

When $s \neq s'$, then

$$\sum_{l=0}^{d-1} e^{i2\pi(s'-s)l/d} \;=\; \sum_{l=0}^{d-1}(e^{i2\pi(s'-s)/d})^l = 0$$

To verify the rightmost of the above inequalites, we notice, that evaluating the geometric sum $\sum_l = 0^{d-1} q^l$ with $q = e^{i2\pi(s'-s)/d}$ leads us to

$$\sum_{l=0}^{d-1} (e^{i2\pi(s'-s)/d})^l = \frac{1 - e^{i2\pi(s-s')}}{1 - e^{(s-s')/d}} = 0.$$

We summarize

- $k \neq j \;\Rightarrow\; r \neq r' \vee s \neq s' \;\Rightarrow \mathrm{tr} v_k^* v_j = 0$
- $k = j \;\Rightarrow\; r = r' \wedge s = s' \;\Rightarrow \mathrm{tr} v_k^* v_j = d$

which shows the first claim. The remaining statements are rather straightforward applications of the first statement and are left as exercises. □

Proof of Theorem 125. Define, using the family $\{v_m\}_{m=1}^{d^2}$ of unitaries from the preceding lemma,

$$\hat{\mathcal{E}}_m(a) \;:=\; v_m^* a v_m \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_A) \qquad\qquad (a \in \mathcal{L}(\mathcal{H})),$$

and

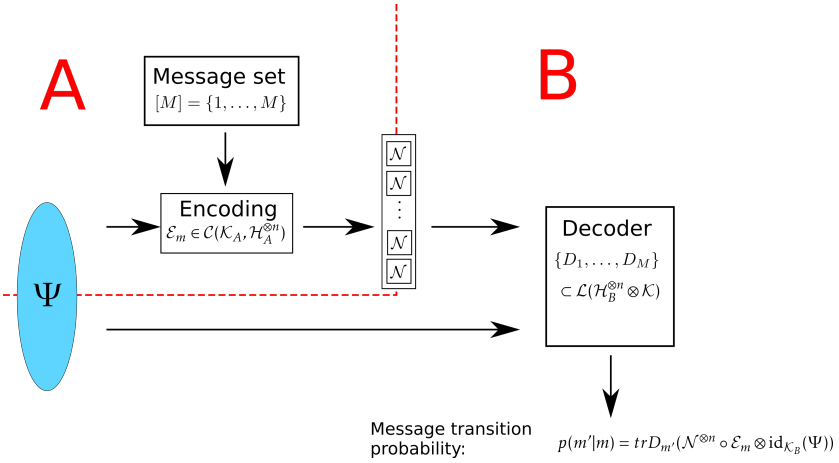
$$\hat{D}_m \;:=\; (\mathbb{1}_{\mathcal{H}_B} \otimes v_m)|\phi\rangle\langle\phi|(\mathbb{1}_{\mathcal{H}_B} \otimes v_m)^* = |\phi_m\rangle\langle\phi_m|$$

where ϕ_m is the maximally entangled vector corresponding to v_m in Lemma 126. Notice, that $\mathcal{E}_m \otimes \mathrm{id}_{\mathcal{H}_B}(|\phi\rangle\langle\phi|) = |\phi_m\rangle\langle\phi_m|$. It is easy to see, that $\{\hat{D}_m\}_{m=1}^{d^2}$ is indeed a POVM. Moreover, it holds

$$\begin{aligned} p(m'|m) &= \mathrm{tr}\left(\mathrm{id} \circ \hat{\mathcal{E}}_m \otimes \hat{D}_{m'} |\phi\rangle\langle\phi|\right) \\ &= \mathrm{tr} \hat{D}_{m'} |\phi_m\rangle\langle\phi_m| \\ &= |\langle\phi_m, \phi_{m'}\rangle|^2 \\ &= \delta_{mm'} \end{aligned}$$

which proves the claim. □

The main goal of this lecture is to give a characterization of the optimal asymptotical classical message transmission rates over a DMQ channel if sender and receiver can choose an arbitrary pure entangled state to assist coding.
The general coding sheme for entanglement-assisted classical message transmission over n uses of a DMQC $\mathcal{N}$ with assistance of a and a shared pure state $\Psi = |\psi\rangle\langle\psi|$ is depicted below.



We aim to determine the optimal asymptotically achievable message transmission rates in the above scenario. First we give the precise definitions. We fix the abbreviation "EA" for "entanglement-assisted".

Definition 127. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a c.p.t.p. map. An* (n, M)-EA message transmission code *for $\mathcal{N}$ is a family $\mathcal{C} := (\Psi, \mathcal{E}_m, D_m)_{m=1}^M$ where*

- *$\Psi := |\psi\rangle\langle\psi| \in \mathcal{S}(\mathcal{K}_A \otimes \mathcal{K}_B)$ is a pure state shared by sender and receiver.*
- *$\mathcal{E}_m \in \mathcal{C}(\mathcal{K}_A, \mathcal{H}_A^{\otimes n})$ is a c.p.t.p. map for each $m \in [M]$, and*
- *$D_m \in \mathcal{L}(\mathcal{H}_B^{\otimes n}, \mathcal{K}_B)$ is a matrix, such that $0 \leq D_m \leq \mathbb{1}_{\mathcal{H}_B^{\otimes n} \otimes \mathcal{K}_B}$, and $\sum_{m=1}^M D_m = \mathbb{1}_{\mathcal{H}_B^{\otimes n} \otimes \mathcal{K}_B}$.*

The *average error* of the code $\mathcal{C}$ is defined by

$$\bar{e}_{EA}(\mathcal{C}, \mathcal{N}^{\otimes n}) := \frac{1}{M} \sum_{m=1}^{M} \mathrm{tr} D_m^c (\mathcal{N}^{\otimes n} \circ \mathcal{E}_m \otimes \mathrm{id}_{\mathcal{K}_B}(\Psi)), \tag{16.1}$$

where we defined, with some abuse of notation $A^c := \mathbb{1} - A^c$ for each matrix A. As we did in case of classical message transmission without entanglement assistance, we define

$$\overline{N}_{EA}(\mathcal{N}, n, \epsilon) := \max\left\{M : \exists\ (n, M)\text{- EA message transm. code } \mathcal{C} \text{ s.t. } \bar{e}_{EA}(\mathcal{C}, \mathcal{N}^{\otimes n}) \leq \epsilon\right\}$$

for each $\epsilon \in [0, 1]$ and $n \in \mathbb{N}$. To state the corresponding capacity theorem, we introduce another quantum entropic quantity.

Definition 128 (Quantum mutual information)**.** *For a c.p.t.p. map $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ and a state $\rho \in \mathcal{S}(\mathcal{H}_A)$, the* quantum mutual information *is defined by*

$$I(\rho, \mathcal{N}) = S(\rho) + S(\mathcal{N}(\rho)) - S(\mathcal{N} \otimes \mathrm{id}_{\mathcal{K}}(|\psi\rangle\langle\psi|)), \tag{16.2}$$

where ψ is the state vector of any purification of ρ.

Remark 129. *Notice, that the term on the r.h.s. of Eq (16.2) above is indeed independent of the choice of purification.*

Theorem 130 (Entanglement-assisted capacity)**.** *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$. It holds*

1. $\forall \epsilon > 0 : \liminf_{n\to\infty} \frac{1}{n} \log \overline{N}_{EA}(\mathcal{N}, n, \epsilon) \geq \sup_{\rho\in\mathcal{S}(\mathcal{H}_A)} I(\rho, \mathcal{N})$, *and*
2. $\inf_{\epsilon>0} : \limsup_{n\to\infty} \frac{1}{n} \log \overline{N}_{EA}(\mathcal{N}, n, \epsilon) \leq \sup_{\rho\in\mathcal{S}(\mathcal{H}_A)} I(\rho, \mathcal{N})$.

Exercise 131. *Convince yourself, that the EA message transmission capacity is the same, when the maximal error criterion is taken into account instead of the average error.*

The claims of Theorem 130 determine the input-state maximized quantum mutual information as the *entanglement-assisted classical capacity* of the QDMC $\mathcal{N}$. Before we prove the claims of the above The following proposition states existence of codes sufficient for proving Theorem 130.1.

Proposition 132. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ be a c.p.t.p. map, and $\sigma \in \mathcal{S}(\mathcal{H}_A)$. For each $\epsilon > 0, \delta > 0$ exists a number n_0 such that for each $n > n_0$*

$$\overline{N}_{EA}(\mathcal{N}, n, \epsilon) \geq \exp\left(n(I(\sigma, \mathcal{N}) - \delta)\right).$$

The strategy to prove the above claim will be, to combine a number of instances of the channel, a given pure bipartite state and certain encoding maps to form an "effective" classical-quantum channel. Afterwards, it will turn out, that classical message transmission codes for this cq channel can be reformulated to give EA message transmission codes for the original channel again. To support this strategy, we need the following lemma.

Lemma 133. *Let $\mathcal{H}$ be a Hilbert space,* $\dim \mathcal{H} := d$, *$\sigma \in \mathcal{S}(\mathcal{H})$, and*

$$\psi = \sum_{i=1}^{d} \sqrt{\alpha_i} v_i \otimes v_i \tag{16.3}$$

be a Schmidt decomposition of a purification ψ of σ ($\alpha_i = 0$ may occur for some i), and $k \in \mathbb{N}$. There is a family

$$\{\tilde{\mathcal{E}}_x\}_{x\in\mathcal{X}} \subset \mathcal{C}(\mathcal{H}^{\otimes k}, \mathcal{H}^{\otimes k}), \tag{16.4}$$

such that for each Hilbert space $\mathcal{K}$, and each $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ with the cq channel $V : \mathcal{X} \to \mathcal{S}(\mathcal{K}^{\otimes n} \otimes \mathcal{K}^{\otimes n})$,

$$V(x) := \mathcal{N}^{\otimes k} \circ \tilde{\mathcal{E}}_x \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k}) \qquad (x \in \mathcal{X}) \tag{16.5}$$

The inequality

$$\left| k \cdot I(\rho, \mathcal{N}) - \chi(q_*, V) \right| \leq 2d \cdot \log(k+1) \tag{16.6}$$

is fulfilled with q_* being the equidistribution on $\mathcal{X}$.

Before proving Lemma 133, we recall some properties of the typical sets. Let $\mathcal{X}$ be an alphabet. For each $k \in \mathbb{N}$, we define the set of *p-typical words of lengths k* by

$$T_p^k := \{x^k = (x_1, \ldots, x_k) : \forall a \in \mathcal{X} : \frac{1}{k} N(a|x^k) = p(a)\},$$

Where $N(a, x^k)$ is the number of occurencies of the letter a in x^k. It is clear, that for some p, k, T_p^k is empty. If T_p^k is nonempty, we call p a *k-type*. We denote the set of all such k-types $\mathcal{T}(\mathcal{X}, k)$. By elementary counting arguments, it holds

$$|\mathcal{T}(\mathcal{X}, k)| \leq (k+1)^{|\mathcal{X}|}, \tag{16.7}$$

i.e. the set of types on a given alphabet does at most increase polynomially with growing blocklengths. Notice, that $T_\lambda^k \cap T_\mu^k = \emptyset$ for all $\lambda \neq \mu \in \mathcal{T}(l, \mathcal{X})$. Moreover,

$$\bigcup_{\lambda \in \mathcal{T}(\mathcal{X},k)} T_\lambda^k = \mathcal{X}^k. \tag{16.8}$$

Summarizing the above relations, we notice that *the collection of sets of typical k-words form a disjoint decomposition of* $\mathcal{X}^k$ The concept of types can be adopted to the quantum setting as follows. Let $\sigma \in \mathcal{S}(\mathcal{H})$ be a given quantum state, and

$$\sigma = \sum_{x \in \mathcal{X}} \alpha_x |\tau_x\rangle\langle\tau_x|$$

be a spectral decomposition of σ with an orthonormal basis $\{\tau_x\}_{x \in \mathcal{X}}$ in $\mathcal{H}$, $|\mathcal{X}| = \dim \mathcal{H}$. We regard α as a probability distribution on $\mathcal{X}$ with $\alpha(x) = \alpha_x$. By the properties of types and sets of typical words, the collection $\{\mathcal{H}_\lambda : \lambda \in \mathcal{T}(\mathcal{X}, k)\}$ with

$$\mathcal{H}_\lambda := span\{\tau_{x^k} := \tau_{x_1} \otimes \cdots \otimes \tau_{x_k} : x^k \in T_\lambda^k\}$$

is a collection of mutually orthogonal subspaces of $\mathcal{H}$ such that $\mathcal{H}^{\otimes k}$ can be written as a direct sum of these spaces, i.e.

$$\mathcal{H}^{\otimes k} = \bigoplus_{\lambda \in \mathcal{T}(\mathcal{X},k)} \mathcal{H}_\lambda.$$

Note, that by construction $\dim \mathcal{H}_\lambda = |T_\lambda^k|$ for each k-type λ.
We are now ready for the proof of Lemma 133.

Proof of Lemma 133. We fix a blocklength k and abbreviate $\mathcal{T}_k := \mathcal{T}(\mathcal{X}, k)$. Using the Schmidt decomposition of ψ from the hypotheses of the lemma, we write

$$\begin{aligned} \psi^{\otimes n} &= \sum_{x^k \in \mathcal{X}^k} \sqrt{\alpha^k(x^k)}\, \tau_{x^k} \otimes \tau_{x^k} \\ &= \sum_{\lambda \in \mathcal{T}} \sum_{x^k \in T_\lambda^k} \sqrt{\alpha^k(x^k)}\, \tau_{x^k} \otimes \tau_{x^k}. \end{aligned}$$

For fixed k-type λ, the α^k-probability is constant over the words in T_λ^k. It holds for each $x^k \in T_\lambda^k$

$$\alpha^k(x^k) = \frac{\alpha^k(T_\lambda^k)}{|T_\lambda^k|} =: \frac{\mu_\lambda}{|T_\lambda^k|}.$$

Therefore, we have with the definition

$$\phi_\lambda := \frac{1}{\sqrt{|T_\lambda^k|}} \sum_{x^k \in T_\lambda^k} \tau_{x^k} \otimes \tau_{x^k}$$

the relations

$$\begin{aligned} \psi^{\otimes k} &= \sum_{\lambda \in \mathcal{T}_k} \sum_{x^k \in T_\lambda^k} \sqrt{\alpha^k(x^k)}\, \tau_{x^k} \otimes \tau_{x^k} \\ &= \sum_{\lambda \in \mathcal{T}_k} \sqrt{\mu_\lambda} \cdot \frac{1}{\sqrt{|T_\lambda^k|}} \sum_{x^k \in T_\lambda^k} \tau_{x^k} \otimes \tau_{x^k} \\ &= \sum_{\lambda \in \mathcal{T}_k} \sqrt{\mu_\lambda} \cdot \phi_\lambda. \end{aligned} \tag{16.9}$$

It is critical to notice here, that for each k-type λ, ϕ_λ is a maximally entangled state vector on $\mathcal{H}_\lambda \otimes \mathcal{H}_\lambda$. For the calculations that follow, we define the shortcuts $d_\lambda := \dim \mathcal{H}_\lambda$ and $\mathcal{Y}_\lambda := [d_\lambda^2] \times \{0,1\}$ for each $\lambda \in \mathcal{T}_k$.
Let for each type λ

$$\{v_{j_\lambda}^\lambda\}_{j_\lambda=1}^{d_\lambda^2} \subset \mathcal{L}(\mathcal{H}_\lambda)$$

be a family of unitaries as stated in Lemma 126. We extend each of these to maps on $\mathcal{L}(\mathcal{H}^{\otimes n})$ via zero-padding, i.e. We define the action of $v_{j_\lambda}^{d_\lambda^2}$ on the orthocomplement of $\mathcal{H}_\lambda$ by

$$v_{j_\lambda}^\lambda x \;:=\; 0 \qquad\qquad (x \in \mathcal{H}_\lambda^\perp) \tag{16.10}$$

Define $\mathcal{Y} = \prod_{\lambda \in \mathcal{T}} \mathcal{Y}_\lambda$ (this is the cartesian product!), and for each $\lambda \in \mathcal{T}_k$, $y_\lambda := (j_\lambda, r_\lambda) \in \mathcal{Y}_\lambda$

$$u_{y_\lambda}^\lambda \;=\; v_{j_\lambda}^\lambda \cdot (-1)^{r_\lambda}.$$

We have, using the representation of $\psi^{\otimes k}$ from Eq. (16.9)

$$(u_{y_\lambda}^\lambda \otimes \mathbb{1}^{\otimes k})\psi^{\otimes k} \;=\; \sum_{\gamma \in \mathcal{T}} \sqrt{\mu_\gamma}(u_{y_\lambda}^\lambda \otimes \mathbb{1}^{\otimes k})\phi_\gamma \;=\; \sqrt{\mu_\lambda}(u_{y_\lambda}^\lambda \otimes \mathbb{1}^{\otimes k})\phi_\lambda. \tag{16.11}$$

The rightmost of the above equalities is by the fact, that $(u_{y_\lambda}^\lambda \otimes \mathbb{1}^{\otimes k})\phi_\gamma = \delta_{\gamma\lambda}\phi_\gamma$. We define for each $y = (y_\lambda)_{\lambda \in \mathcal{T}_k} \in \mathcal{Y}$

$$u_y \;:=\; \sum_{\lambda \in \mathcal{T}_k} u_{y_\lambda}^\lambda,$$

and $\tilde{\mathcal{E}}_y(a) \;:=\; u_y(a)u_y^*$ for each $y \in \mathcal{Y}$. Therefore, we have for all $y = (y_\lambda)_{\lambda \in \mathcal{T}_k}$

$$\begin{aligned} \tilde{\mathcal{E}}_y \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k}) &= \sum_{\lambda,\gamma \in \mathcal{T}_k} (u_{y_\lambda}^\lambda \otimes \mathbb{1}^{\otimes k})|\psi\rangle\langle\psi|^{\otimes k}(u_{y_\gamma}^\gamma \otimes \mathbb{1}^{\otimes k})^* \\ &= \sum_{\lambda,\gamma \in \mathcal{T}_k} (u_{y_\lambda}^\lambda \otimes \mathbb{1}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\gamma|(u_{y_\gamma}^\gamma \otimes \mathbb{1}^{\otimes k})^*. \end{aligned}$$

To proceed with the proof we need to show two identities.

$$\text{Identity 1:} \qquad \frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}}\tilde{\mathcal{E}}_y\otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k}) \;=\; \sum_{\lambda\in\mathcal{I}}\mu_\lambda\pi_\lambda\otimes\pi_\lambda,$$

where $\pi_\lambda = \frac{\mathbb{1}_{\mathcal{H}_\lambda}}{d_\lambda}$ is the maximally mixed state on $\mathcal{H}_\lambda$ for each type λ. To prove the above equation, we consider the sum

$$\begin{aligned}\frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}}(\tilde{\mathcal{E}}_y\otimes \mathrm{id}_{\mathcal{H}}^{\otimes k})(|\psi\rangle\langle\psi|^{\otimes k}) &= \frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}}\sum_{\lambda,\gamma\in\mathcal{I}_k}\sqrt{\mu_\lambda\cdot\mu_\gamma}(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\psi\rangle\langle\psi|^{\otimes k}(u_{y_\gamma}^\gamma\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\\ &= \frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}}\sum_{\lambda,\gamma\in\mathcal{I}_k}\sqrt{\mu_\lambda\cdot\mu_\gamma}(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\gamma|(u_{y_\gamma}^\gamma\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\end{aligned}$$

We evaluate the inner sums in the last line above for two cases.

- $\lambda\neq\gamma$: We have

  $$\begin{aligned}&\sum_{y\in\mathcal{Y}}\sqrt{\mu_\lambda\cdot\mu_\gamma}(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\gamma|(u_{y_\gamma}^\gamma\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\\ &= \left(\frac{|\mathcal{Y}|}{|\mathcal{Y}_\lambda|}\right)^2\sum_{y_\lambda\in\mathcal{Y}_\lambda}\sum_{y_\gamma\in\mathcal{Y}_\gamma}\sqrt{\mu_\lambda\cdot\mu_\gamma}(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\gamma|(u_{y_\gamma}^\gamma\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\\ &= \left(\frac{|\mathcal{Y}|}{|\mathcal{Y}_\lambda|}\right)^2\sum_{r_\lambda,r_\gamma\in\{0,1\}}(-1)^{r_\lambda+r_\gamma}\sum_{j_\lambda\in[d_\lambda^2]}\sum_{j_\gamma\in[d_\gamma^2]}\sqrt{\mu_\lambda\cdot\mu_\gamma}(v_{j_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\gamma|(v_{j_\gamma}^\gamma\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\\ &= 0\end{aligned}$$

  The last inequality above is by the fact, that $\sum_{r_\lambda,r_\gamma\in\{0,1\}}(-1)^{r_\lambda+r_\gamma}$ vanishes.

- $\lambda=\gamma$. In this case, The inner sum reads

  $$\begin{aligned}&\sum_{y\in\mathcal{Y}}\mu_\lambda(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\lambda|(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\\ &= \frac{|\mathcal{Y}|}{|\mathcal{Y}_\lambda|}\sum_{y_\lambda\in\mathcal{Y}_\lambda}\mu_\lambda(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\lambda|(u_{y_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\\ &= \frac{|\mathcal{Y}|}{|\mathcal{Y}_\lambda|}\sum_{r_\lambda}(-1)^{2\cdot r_\lambda}\sum_{j_\lambda\in[d_\lambda^2]}\mu_\lambda(v_{j_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})|\phi_\lambda\rangle\langle\phi_\lambda|(v_{j_\lambda}^\lambda\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k})^*\\ &= \frac{|\mathcal{Y}|}{|\mathcal{Y}_\lambda|}2\mu_\lambda\cdot\mathbb{1}_{\mathcal{H}}^{\otimes k}\otimes\mathbb{1}_{\mathcal{H}}^{\otimes k}\\ &= \frac{|\mathcal{Y}|}{|\mathcal{Y}_\lambda|}2d_\lambda^2\mu_\lambda\pi_\lambda\otimes\pi_\lambda\\ &= |\mathcal{Y}|\mu_\lambda\pi_\lambda\otimes\pi_\lambda.\end{aligned}$$

The second identity is

$$\text{Identity 2:} \qquad (\mathcal{N}\circ\tilde{\mathcal{E}}_y\otimes\mathrm{id}_{\mathcal{H}}^{\otimes k})(|\psi\rangle\langle\psi|^{\otimes k}) \;=\; (\mathbb{1}^{\otimes k}\otimes u_y^T)\mathcal{N}\otimes\mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k})(\mathbb{1}^{\otimes k}\otimes u_y^T)^*$$

holding for any given c.p.t.p. map $\mathcal{N} \in \mathcal{C}(\mathcal{H}^{\otimes k}, \mathcal{K})$ with a Hilbert space $\mathcal{K}$. To prove Identity 2, we remember the claim of Lemma 49 from Chapter 5. If ϕ is the state vector of a maximally entangled pure state on $\mathcal{F} \otimes \mathcal{F}$, and A a matrix on $\mathcal{F}$, then

$$(\mathbb{1}_{\mathcal{F}} \otimes A)\phi \;=\; (A^T \otimes \mathbb{1}_{\mathcal{F}})\phi.$$

This fact applied on each $\lambda \in \mathcal{T}_k$ together with the representation in (16.9) proves the claim.
From Identity 2 and unitary invariance of the von Neumann entropy, we directly obtain

$$S(\mathcal{N} \circ \tilde{\mathcal{E}}_y \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k})) \;=\; S(\mathcal{N} \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k})) \tag{16.12}$$

Having these equalities, we can calculate

$$\begin{aligned}
S\left(\frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}} \mathcal{N}^{\otimes k} \circ \tilde{\mathcal{E}}_y \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k})\right) \;&=\; S\left(\mathcal{N}^{\otimes k} \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}\left(\sum_{\lambda\in\mathcal{T}} \mu_\lambda \pi_\lambda \otimes \pi_\lambda\right)\right) \\
&\geq \sum_{\lambda\in\mathcal{T}_k} \mu_\lambda S(\mathcal{N}^{\otimes k} \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(\pi_\lambda \otimes \pi_\lambda)) \\
&= \sum_{\lambda\in\mathcal{T}_\lambda} \mu_\lambda \Big(S(\mathcal{N}^{\otimes k}(\pi_k)) + S(\pi_\lambda)\Big) \\
&\geq S\left(\mathcal{N}^{\otimes k}\left(\sum_{\lambda\in\mathcal{T}_k} \mu_\lambda \pi_\lambda\right)\right) \;+\; S\left(\sum_{\lambda\in\mathcal{T}_k} \mu_\lambda \pi_\lambda\right) - 2\, H(\mu) \\
&\geq S(\mathcal{N}(\sigma)^{\otimes k}) \;+\; S(\sigma^{\otimes k}) - d \cdot \log(k+1)
\end{aligned}$$

The first inequality above is by concavity, the second by almost-convexity of the von Neumann entropy (Lemma **??**). The last inequality is by our initial representation of $\sigma^{\otimes k}$ in terms of mutually orthogonal projections onto typical subspaces together with the bound $H(\mu) \leq \log|\mathcal{T}_k| \leq d \cdot \log(k+1)$.
Choosing q_* and V as in the hypotheses of the proposition, we have

$$\begin{aligned}
\chi(q_*, V) \;&=\; S\left(\frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}} \mathcal{N}\otimes k \circ \tilde{\mathcal{E}}_y \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k})\right) - \frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}} S(\mathcal{N}^{\otimes k} \circ \tilde{\mathcal{E}}_y \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k})) \\
&\geq\; S(\mathcal{N}(\sigma)^{\otimes k}) \;+\; S(\sigma^{\otimes k}) - d \cdot \log(k+1) - S(\mathcal{N}^{\otimes k} \otimes \mathrm{id}_{\mathcal{H}}^{\otimes k}(|\psi\rangle\langle\psi|^{\otimes k})) \\
&=\; k \cdot I(\sigma, \mathcal{N}) \;-\; d \cdot \log(k+1)
\end{aligned}$$

The reversed inequality can be proven in a similar way. □

Now we are ready to prove Proposition 132.

Proof of Proposition 132. Set $d_A := \dim \mathcal{H}_A$, $d_B := \dim \mathcal{H}_B$, and fix a state $\sigma \in \mathcal{S}(\mathcal{H}_A)$. Let

$$\sigma \;=\; \sum_{i=1}^{d_A} \alpha_i \tau_i$$

be a spectral decomposition of σ. Define a pure state $\Psi := |\psi\rangle\langle\psi| \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_A)$ via

$$\psi := \sum_{i=1}^{d_A} \sqrt{\alpha_i}\tau_i \otimes \tau_i.$$

Let $\{\tilde{\mathcal{E}}\}_{x\in\mathcal{X}}$ be a family of channels as in Lemma 133. Define the cq channel $V : \mathcal{X} \to \mathcal{S}(\mathcal{H}_B^{\otimes k} \otimes \mathcal{H}_B^{\otimes k})$ by

$$V(x) \;:=\; \mathcal{N}^{\otimes k} \circ \tilde{\mathcal{E}}_x \otimes \mathrm{id}_{\mathcal{H}_B}^{\otimes k}(\Psi^{\otimes k}) \qquad\qquad (x \in \mathcal{X}).$$

With Lemma 133, we know, that

$$\chi(q_*, V) \;\geq\; kI(\sigma, \mathcal{N}) - d_A \log(k+1)$$

holds.
Fix $\epsilon, \delta > 0$, and an arbitrary blocklength $n \in \mathbb{N}$. Write $n = a \cdot k + b$ with $a, b \in \mathbb{N}$, $0 \leq b < k$. If a is large enough (which happens for large enough n), we have

$$N(a, V, \epsilon) \;\geq \exp(a(\chi(q_*, V) - \frac{\delta}{2})).$$

This follows from the achievability statement in Holevo's Theorem. Consequently, there is an (a, M) code $\tilde{\mathcal{C}} = (u_m, \tilde{D}_m)_{m=1}^M$ for classical message transmission over the DMCQ channel V with

$$M \;\geq\; \exp\Big(a(\chi(q_*, V) - \frac{\delta}{2})\Big)$$

and average transmission error $\overline{e}(\tilde{\mathcal{C}}, V^{\otimes a}) \;\leq\; \epsilon$. We will now use $\tilde{\mathcal{C}}$ to construct an (n, M) code for entanglement-assisted message transmission over $\mathcal{N}$.
Define, for each $m \in [M]$, and each codeword $u_m = (u_{m,1}, \ldots, u_{m,a})$ from $\tilde{\mathcal{C}}$ a c.p.t.p. map $\mathcal{E}_m \in \mathcal{C}(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_A^{\otimes n})$ by

$$\mathcal{E}_m(\cdot) \;:=\; \tilde{\mathcal{E}}_{u_{m,1}} \otimes \cdots \otimes \mathcal{E}_{u_{m,a}} \otimes \mathrm{id}_{\mathcal{H}}^{\otimes b}(\cdot),$$

and a POVM $\{D_m\}_{m=1}^M$ on $(\mathcal{H}_B \otimes \mathcal{H}_B)^{\otimes n}$ by

$$D_m \;:=\; \tilde{D}_m \otimes \mathbb{1}_{\mathcal{H}_B \otimes \mathcal{H}_B}^{\otimes b}.$$

With the preceding definitions, and the pure maximally entangled state $\Psi := |\psi\rangle\langle\psi|$, $\mathcal{C} := (\Psi, \mathcal{E}_m, D_m)_{m=1}^M$ is an (n, M)-EA message transmission code for $\mathcal{N}$. Moreover, it holds for each $m \in [M]$

$$\begin{aligned}\mathrm{tr} D_m\Big(\mathcal{N}^{\otimes n} \circ \mathcal{E}_m(\Psi^{\otimes n})\Big) \;&=\; \mathrm{tr}\tilde{D}_m\Big(\bigotimes_{i=1}^{a} \mathcal{N}^{\otimes k} \circ \tilde{\mathcal{E}}_{u_{m,i}} \otimes \mathrm{id}_{\mathcal{H}_B}^{\otimes k}(\Psi^{\otimes k})\Big) \cdot \mathrm{tr}(\mathcal{N}^{\otimes b} \otimes \mathrm{id}_{\mathcal{H}_B}^{\otimes b}(\Psi^{\otimes b})) \\ &= \mathrm{tr}\tilde{D}_m V^{\otimes a}(u_m).\end{aligned}$$

Arithmetic averaging of the above equality over all messages leads us to

$$\overline{e}_{EA}(\mathcal{C}, \mathcal{N}^{\otimes n}) \;=\; 1 - \frac{1}{M}\sum_{m=1}^{M} \mathrm{tr}\tilde{D}_m V^{\otimes a}(u_m) \;\leq \epsilon$$

Therefore,

$$\begin{aligned} N(n,\mathcal{N},\epsilon) \;&\geq\; M \\ &\geq \exp\Big(a(\chi(q_*,V)-\frac{\delta}{2})\Big) \\ &> \exp\Big(n\Big(\frac{1}{k}\chi(q_*,V)-\frac{\delta}{2k}-\frac{1}{kn}(\chi(q_*,V)\cdot\frac{\delta}{2k})\Big)\Big) \end{aligned} \tag{16.13}$$

where the last inequality stems from

$$a=\frac{n-b}{k}>\frac{n-k}{k}=\frac{n}{k(1-\frac{1}{n})}.$$

Evaluating the exponent on the right hand side of the inequality in (16.13), we obtain using Lemma 133

$$\frac{1}{k}\chi(q_*,V)-\frac{\delta}{2k}-\frac{1}{kn}(\chi(q_*,V)\cdot\frac{\delta}{2k}) \;\geq I(\sigma,\mathcal{N})-\frac{1}{k}(d_A\log(k+1))-\frac{\delta}{2}-\frac{1}{n}\chi(q_*,V)-\frac{\delta}{2n}. \tag{16.14}$$

If n is large enough, we can infer from (16.13) and (16.14), that the claimed inequality

$$\overline{N}_{EA}(n,\mathcal{N},\epsilon) \;\geq\; \exp(n(I(\sigma,\mathcal{N})-\delta))$$

is fulfilled. □

Lemma 134. *Let $\mathcal{N}\in\mathcal{C}(\mathcal{H}_A,\mathcal{H}_B)$, $\Psi\in\mathcal{S}(\mathcal{H}_A\otimes\mathcal{K}_B)$ be a pure state, $\{\mathcal{E}_m\}_{m=1}^M\subset\mathcal{C}(\mathcal{H}_A,\mathcal{H}_A)$, and $q\in\mathcal{P}([M])$. Define the c.p.t.p. map $\mathcal{E}(\cdot):=\sum_{m\in[M]}q(m)\mathcal{E}_m(\cdot)$, and $\rho_A = \mathrm{tr}_{\mathcal{K}_B}\Psi$. With the cq channel $V:\,[M]\to\mathcal{S}(\mathcal{H}_B\otimes\mathcal{K}_B)$,*

$$m \;\mapsto\; V(m) \;:=\; \mathcal{N}\circ\mathcal{E}_m\otimes\mathrm{id}_{\mathcal{K}_B}(\Psi),$$

it holds

$$\chi(q,V) \;\leq\; I(\mathcal{E}(\rho_A),\mathcal{N}). \tag{16.15}$$

Proof. We introduce the shortcuts

$$\tilde{\rho}_m \;:=\; \mathcal{E}_m\otimes\mathrm{id}_{\mathcal{K}_B}(\Psi), \quad \text{and} \quad \rho_B \;:=\; \mathrm{tr}_{\mathcal{K}_A\Psi}$$

for each $m\in[M]$. By subadditivity of the von Neumann entropy, we have

$$\begin{aligned} \chi(q,V) \;&=\; S\Big(\mathcal{N}\circ\mathcal{E}\otimes\mathrm{id}_{\mathcal{K}_B}(\Psi)\Big)-\sum_{m=1}^M q(m)\,S(\mathcal{N}\otimes\mathrm{id}_{\mathcal{K}_B}(\tilde{\rho}_m)) \\ &\leq S\left(\mathcal{N}\circ\mathcal{E}(\rho_A)\right)+S(\rho_B)-\sum_{m=1}^M q(m)\,S(\mathcal{N}\otimes\mathrm{id}_{\mathcal{K}_B}(\tilde{\rho}_m)) \end{aligned}$$

We aim to further bound the expression

$$\sum_{m=1}^M q(m)\Big(S(\rho_B)-S(\mathcal{N}\otimes\mathrm{id}_{\mathcal{K}_B}(\tilde{\rho}_m))\Big). \tag{16.16}$$

To this reason, we first derive an estimate for each of the summands on the r.h.s. of the above inequality. Let, with a suitable additional Hilbert space $\mathcal{H}_R$, for each $m \in [M]$ $\Gamma_m \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_R)$ be a purification or $\tilde{\rho}_{A,m}$. We show the inequality

$$S(\rho_B) - S(\mathcal{N} \otimes \mathrm{id}_{\mathcal{K}_B}(\tilde{\rho}_m)) \;\leq\; S(\tilde{\rho}_{A,m}) - S(\mathcal{N} \otimes \mathrm{id}_{\mathcal{H}_R}(\Gamma_m)). \tag{16.17}$$

Let $u_m : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ be a Stinespring isometry for $\mathcal{E}_m$, and $\mathcal{U}_m(\cdot) := u_m(\cdot)u_m^*$ the corresponding unitary transformation. Observe, that $\mathcal{U}_m \otimes \mathrm{id}_{\mathcal{K}_B}(\Psi)$ is a purification of $\tilde{\rho}_{A,m}$. Consequently

$$S\left((\mathcal{N} \otimes \mathrm{id}_{\mathcal{H}_E \otimes \mathcal{K}_B}) \circ (\mathcal{U}_m \otimes \mathrm{id}_{\mathcal{K}_B})\Psi\right) \;=\; S((\mathcal{N} \otimes \mathrm{id}_{\mathcal{H}_R})(\Gamma_m))$$

holds. Moreover,

$$S(\tilde{\rho}_{A,m}) \;=\; S(\mathrm{tr}_{\mathcal{H}_A}\Gamma_m) \;=\; S(\mathrm{tr}_{\mathcal{H}_A}(\mathcal{U}_m \otimes \mathrm{id}_{\mathcal{H}_B})(\Psi))$$

The rightmost inequality above holds, because Γ_m and $\mathcal{U} \otimes \mathrm{id}_{\mathcal{H}_B}(\Psi)$ are both purifications of $\tilde{\rho}_{A,m}$. For the calculations to come, we use the state

$$\sigma_m \;:=\; (\mathcal{N} \otimes \mathrm{id}_{\mathcal{K}_B \otimes \mathcal{H}_E}) \circ (\mathcal{U}_m \otimes \mathrm{id}_{\mathcal{K}_B})(\Psi)$$

It then holds

$$\begin{aligned} S(\rho_B) - S(\mathcal{N} \otimes \mathrm{id}_{\mathcal{K}_B}(\tilde{\rho}_m)) \;&=\; S(\sigma_{B,m}) - S(\sigma_{BB',m}) \\ &=\; -S(B'|B, \sigma_{BB'E,m}) \\ &\leq\; -S(B'|BE, \sigma_{BB'E,m}) \\ &=\; S(\sigma_{BE,m}) - S(\sigma_{BB'E,m}) \\ &=\; S(\mathrm{tr}_{\mathcal{H}_A}\Gamma_m) - S(\mathcal{N} \otimes id(\Gamma_m)). \end{aligned}$$

Therefore, (16.17) is valid. We will now show, that the function $f : \mathcal{S}(\mathcal{H}_A) \to \mathbb{R}$,

$$f(\tau) \;:=\; S(\tau) - S(\mathcal{N} \otimes \mathrm{id}_{\mathcal{H}_R}(\Phi))$$

is concave. Introduce the isometric channel $\mathcal{V}(\cdot) := v(\cdot)v^*$ where $v : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ is a Stinespring isometry for $\mathcal{N}$ with a suitable additional Hilbert space $\mathcal{H}_E$. Define

$$\tilde{\Phi} \;:=\; \mathcal{V} \otimes \mathrm{id}_{\mathcal{H}_R}(\Phi).$$

(notice, that the state defined is pure). Denote the marginals of $\tilde{\Phi}$ on the respective subsystems by $\gamma_A, \gamma_{BE}, \gamma_{BR}$. It holds

$$S(\tau) = S(\gamma_{BE}), \qquad S(\gamma_{BR}) = S(\gamma_E).$$

Therfore

$$\begin{aligned} f(\tau) \;=\; S(\tau) - S(\mathcal{N} \otimes \mathrm{id}_{\mathcal{H}_R}(\Phi)) \;&=\; S(\gamma_{BE}) - S(\gamma_E) \\ &=\; S(B|E, \gamma_{BE}) \\ &=\; S(B|E, \mathcal{V}(\tau)) \end{aligned}$$

Since the map $\tau \mapsto \mathcal{V}(\tau)$ is affine, and the conditional von Neumann entropy is concave, f is indeed a concave function. Putting everything together, we have

$$\begin{aligned}
\chi(q,V) &\leq S(\mathcal{E}(\rho_A)) + \sum_{m=1}^{M} q(m)\Big(S(\rho_B) - S(\mathcal{N}\otimes \mathrm{id}_{\mathcal{K}_B}(\tilde{\rho}_m))\Big) \\
&\leq S(\mathcal{E}(\rho_A)) + \sum_{m=1}^{M} q(m)\Big(S(\mathrm{tr}_{\mathcal{H}_A}\Gamma_m) - S(\mathcal{N}\otimes \mathrm{id}_{\mathcal{K}_B}(\Gamma_m))\Big) \\
&= S(\mathcal{E}(\rho_A)) + \sum_{m=1}^{M} q(m) f(\tilde{\rho}_{A,m}) \\
&\leq S(\mathcal{E}(\rho_A)) + f(\mathcal{E}(\rho_A)) \\
&\leq I\left(S(\mathcal{E}(\rho_A)), \mathcal{N}\right).
\end{aligned}$$

□

We are now equipped with the prerequisites necessary to prove Theorem 130.

Proof of Theorem 130. The achievability part (Theorem 130.1) follows indeed from Proposition 132 via executing the lower limit. To prove the converse (Theorem 130.2), fix an arbitrary blocklength n and let $\mathcal{C} := (\Phi, \mathcal{E}_m, D_m)_{m=1}^{M}$ be an (n,M)-EA message transmission code with

$$\overline{e} := \overline{e}_{EA}(\mathcal{C}, \mathcal{N}^{\otimes n}) < 1.$$

We define states

$$\rho = \mathrm{tr}_{\mathcal{K}_B}\Phi, \quad \text{and} \quad \overline{\tau} := \sum_{m=1}^{M} p_*(m)\mathcal{E}_m(\rho). \tag{16.18}$$

where p_* denotes the equidistribution on the message set, i.e. $p_*(m) = \frac{1}{M}$ for each m. We define $\overline{\tau}_i$ as the marginal state deriving from $\overline{\tau}$ on the ith tensor factor of $\mathcal{H}^{\otimes n}$. By subadditivity of the quantum mutual information, it holds

$$I(\overline{\tau}, \mathcal{N}^{\otimes n}) \leq \sum_{i=1}^{n} I(\overline{\tau}_i, \mathcal{N}). \tag{16.19}$$

Define a cq channel $V : [M] \to \mathcal{S}(\mathcal{H}_B^{\otimes n} \otimes \mathcal{K}_B)$ by

$$V(m) := (\mathcal{N}^{\otimes n} \circ \mathcal{E}_m \otimes \mathrm{id}_{\mathcal{K}_B})(\Phi).$$

Let X be a random variable with values in $[M]$ and

$$\Pr(X = m) = \frac{1}{M} = p_*(m), \tag{16.20}$$

and conditional probabilities

$$\Pr(Y = m' | X = m) := \mathrm{tr} D_{m'} V(m) \qquad (m, m' \in [M])$$

Then

$$\Pr(X \neq Y) \;=\; \overline{e}_{EA} := \overline{e}.$$

Using the above inequality and Fano's Lemma, we have

$$\begin{aligned} H(X|Y) &\leq \Pr(X \neq Y)\log M + h(\Pr(X \neq Y)) \\ &= \overline{e}\log M + h(\overline{e}) \\ &\leq \overline{e}\log M + 1. \end{aligned} \tag{16.21}$$

Moreover, the following chain of (in)equalities is valid.

$$\chi(p_*, V) \;\leq\; I(\overline{\tau}, \mathcal{N}^{\otimes n}) \;\leq \sum_{i=1}^{n} I(\overline{\tau}_i, \mathcal{N}) \;\leq n \cdot I\left(\frac{1}{n}\sum_{i=1}^{n}\overline{\tau}_i, \mathcal{N}\right) \leq n \sup_{\rho \in \mathcal{S}(\mathcal{H}_A)} I(\rho, \mathcal{N}). \tag{16.22}$$

The first of the above inequalities is justified by Lemma 134, and the definition of $\overline{\tau}$. The second is Eq. (16.19). The third is by concavity of the quantum mutual information in the first argument. It consequently holds

$$\begin{aligned} \log M \;&=\; H(X) \\ &=\; I(X;Y) + H(X|Y) \\ &\leq\; I(X;Y) + \overline{e}\log M + 1 \\ &\leq\; \chi(p_*, V) + \overline{e}\log M + 1 \\ &\leq n \sup_{\rho \in \mathcal{S}(\mathcal{H}_A)} I(\rho, \mathcal{N}) + \overline{e}\log M + 1. \end{aligned}$$

The first inequality above is (16.21). The second is by Holevo's bound, and the third is (16.22). Since $\mathcal{C}$ was arbitrary,

$$\frac{1}{n}\log \overline{N}_{EA}(n, \mathcal{N}, \overline{e}) \;\leq \sup_{\rho \in \mathcal{S}(\mathcal{H}_A)} I(\rho, \mathcal{N}) + \frac{\overline{e}\log M}{n} + \frac{1}{n}. \tag{16.23}$$

Taking limits on both sides of the above inequality proves the converse. □

Part II

Supplements

Mathematical preliminaries and notation

In this lecture we review some basic mathematical definitions which may in most parts be taught in usual higher math courses. At the same time we fix the notation for the forthcoming lectures, and get used to the so-called Dirac notation common in quantum (information) theory.

16.1 Linear algebra

The mathematical playground for quantum theory are *Hilbert spaces* (i.e. a linear space with a scalar product, which is closed in the norm deriving from the scalar product.) For this course, we will assume, that each Hilbert space is a finite dimensional euclidean space over the field of complex numbers. Under this restriction, the terminology "Hilbert space" is rather superfluous. However, it is standard in the quantum information theory literature also for finite dimensions, therefore, we keep it.
If we fix a basis for an Hilbert space $\mathcal{H}$, $d := \dim \mathcal{H} < \infty$, $\mathcal{H}$ is isomorph to $\mathbb{C}^d$, where each vector $v \in \mathcal{H}$ corresponds to a column vector

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix} \tag{16.24}$$

with entries $v_1, \ldots, v_d \in \mathbb{C}$. We use the standard euclidean scalar product $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ defined by

$$\langle v, w \rangle := \sum_{i=1}^{d} \overline{v}_i w_i \tag{16.25}$$

for each $v = (v_1, \ldots, v_d),\ w = (w_1, \ldots, w_d) \in \mathbb{C}^d$ (where $\overline{v}$ is the notation for the complex conjugate of v.)
We will freely switch between the abstract and component notation, i.e. assuming that an orthonormal basis is fixed, we will not distinguish between a d-dimensional Hilbert space $\mathcal{H}$ and $\mathbb{C}^d$. If not otherwise stated, we assume the basis to be the canonical orthonormal basis $\{e_i\}_{i=1}^d$ where

$$e_i := \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \tag{16.26}$$

with the i-th entry being 1 and all other entries being 0. With two Hilbert spaces $\mathcal{H}$, and $\mathcal{H}'$ we denote the set of linear maps from $\mathcal{H}$ to $\mathcal{H}'$ by $\mathcal{L}(\mathcal{H},\mathcal{H}')$. If we fix a bases in the underlying spaces $\mathcal{H}$, $\mathcal{H}'$, $\mathcal{L}(\mathcal{H},\mathcal{H}')$ is isomorphic to the set $\mathbb{M}_{d\times d'}(\mathbb{C})$ of $d\times d'$ matrices with complex entries, sometimes, we will also write a matrix A as the collection of its entries. $A=(a_{ij})_{i=1,j=1}^{m,n}$ then corresponds to the matrix

$$A:=\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \tag{16.27}$$

If $\mathcal{H}'$ equals $\mathcal{H}$, we will also use the shortcut $\mathcal{L}(\mathcal{H})$ for the set $\mathcal{L}(\mathcal{H},\mathcal{H}')$. We denote the *adjoint matrix* to A by A^*, the *transposed matrix* to A by A^T, i.e.

$$A^*:=\begin{pmatrix} \overline{a}_{11} & \cdots & \overline{a}_{m1} \\ \vdots & \ddots & \vdots \\ \overline{a}_{1n} & \cdots & \overline{a}_{nm} \end{pmatrix} \quad \text{and} \quad A^T:=\begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nm} \end{pmatrix}. \tag{16.28}$$

The reader should note, that the transposition is dependent on the chosen basis, while the adjoint is not. The *trace* of a square matrix $A\in\mathcal{L}(\mathcal{H})$ is defined by

$$\mathrm{tr}(A):=\sum_{v_i,Av_i}, \tag{16.29}$$

where $\{v_1,\ldots,v_{\dim_{\mathcal{H}}}\}$ is an orthonormalbasis in $\mathcal{H}$. Using the trace, we can define the *Hilbert-Schmidt scalar product* on $\mathcal{L}(\mathcal{H})$,

$$\langle A,B\rangle_{HS} := \mathrm{tr}(A^*B) \qquad (A,B\in\mathcal{L}(\mathcal{H})). \tag{16.30}$$

By the (finite dimensional) Riesz representation theorem, there is a one-to-one relationship between $\mathcal{H}$ and its dual space $\mathcal{H}^*$, i.e. to each linear functional $f\in\mathcal{H}^*$ there is a unique element $v_f\in\mathcal{H}$, such that

$$f(w):=\langle v_f,w\rangle \tag{16.31}$$

for each $w\in\mathcal{H}$. This fact is reflected by the so-called **Dirac notation**. In this notation, each element $v\in\mathcal{H}$ is written as a "ket" $|v\rangle$, while the corresponding element $v^*\in\mathcal{H}^*$ with

$$v^*(w)=\langle v,w\rangle \tag{16.32}$$

for each $w\in\mathcal{H}$ is written as a "bra" $\langle v|$. Note, that $\langle\alpha\cdot v|=\overline{\alpha}\langle v|$ for all $\alpha\in\mathbb{C}$. Moreover, $\langle v_1+v_2|=\langle v_1|+\langle v_2|$. The *outer product* of $w\in\mathcal{H},\in\mathcal{H}'$, $|v\rangle\langle w|\in\mathcal{L}(\mathcal{H},\mathcal{H}')$ is the rank one matrix with the property, that

$$(|v\rangle\langle w|)|x\rangle=\langle w,x\rangle|v\rangle \tag{16.33}$$

holds for all for all $x\in\mathcal{H}$. It holds $(|v\rangle\langle w|)^*=\langle w||v\rangle$. The canonical basis in $\mathcal{L}(\mathcal{H},\mathcal{H}')$ is given by the *matrix units* $\{E_{ij}\}_{i=1,j=1}^{\dim\mathcal{H},\mathcal{H}'}$, where E_{ij} is the matrix with the i,j-entry

being one and all others being zero. The following subsets of $\mathcal{L}(\mathcal{H})$ will be of some importance for our considerations.

$$\mathcal{L}^h(\mathcal{H}) := \{A \in \mathcal{L}(\mathcal{H}) : \ A^* = A\} \qquad \text{(Hermitian maps)}$$
$$\mathcal{L}^+(\mathcal{H}) := \{A \in \mathcal{L}(\mathcal{H}) : \ \forall x \in \mathcal{H} : \ \langle x, Ax\rangle \geq 0\} \qquad \text{(positive semidefinite maps)}$$

We call a map $p \in \mathcal{L}(\mathcal{H})$ an *(orthogonal) projection* , if it is hermitian and idempotent, i.e. $p = pp = p^2$.
We call a number $\lambda \in \mathbb{C}$ an eigenvalue of $A \in \mathbb{M}_n$, if there exists $x \in \mathbb{C}^n$, $x \neq 0$, such that $Ax = \lambda x$ (in wich case x is called an eigenvector for A to eigenvalue λ, equivalently, λ is an eigenvalue. We denote the spectrum of A (the set of eigenvalues) by $\mathrm{spec}(A)$. We say two maps $A, B \in \mathbb{M}_n$ commute, if their commutator , i.e. the function

$$[A, B] := AB - BA \tag{16.34}$$

vanishes. A very important consequence of being hermitian is existence of an orthonormal basis of eigenvectors, i.e. a spectral decomposition.

Theorem 135 (Spectral decomposition)**.** *Let $A \in \mathcal{L}(\mathcal{H})$ be normal. Then there exists an orthonormal basis $\{v_i : \ 1 \leq i \leq \dim \mathcal{H}\}$, such that*

$$\sum_{i=1}^{\dim \mathcal{H}} \alpha_i |v_i\rangle\langle v_i| \tag{16.35}$$

holds. The numbers $\alpha_1, \cdots, \alpha_{\dim \mathcal{H}}$ are the eigenvalues of A (counted with multiplicities).

If $A \in \mathcal{L}^h(\mathcal{H})$, then all eigenvalues are real numbers.

Theorem 136 (Singular value decomposition)**.** *Let $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$. Then there exist orthonormal systems $\{\varphi\}_{i=1}^r \subset \mathcal{H}$ and $\{\psi\}_{j=1}^r \subset \mathcal{K}$ such that*

$$A = \sum_{i,j=1}^{r} \sigma_j(A) |\varphi_i\rangle\langle\psi_j| \tag{16.36}$$

holds, where r is the rank of A and $\sigma_1(A) \geq \cdots \geq \sigma_r(A) > 0$ are the singular values (i.e. the eigenvalues of the positive semidefinite matrix A^*A.)

The spectral theorem allows to exted real functions in a natural way to hermitian matrices. Let $\mathcal{W} \subset \mathbb{R}$, and $f : \mathcal{W} \to \mathbb{R}$. We define for each hermitian map $A \in \mathcal{L}(\mathcal{H})$ with $\mathrm{spec}(A) \subset \mathcal{W}$

$$f(A) := \sum_{i=1}^{\dim \mathcal{H}} f(\alpha_i) |v_i\rangle\langle v_i|, \tag{16.37}$$

using a spectral decomposition $A := \sum_{i=1}^{\dim \mathcal{H}} \alpha_i |v_i\rangle\langle v_i|$. Particular examples of such matrix-functions we consider in this lecture are the square-root, the exponential functions and logarithms.

Remark 137. *Using matrix functions one has to be very cautious when applying properties of the real functions in the matrix setting. E.g. the identities*

$$\sqrt{AB} = \sqrt{A}\sqrt{B}, \text{ and } \exp A + B = \exp A \exp B \tag{16.38}$$

do not hold in general. We also can define a matrix version of the absolute value of a complex number.

Theorem 138 (Polar decomposition)**.** *Let $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$. We can write A in the form*

$$A := |A|U \tag{16.39}$$

where $|A| := \sqrt{A^*A}$, and U is a (partial) isometry.

Important norms on $\mathcal{L}(\mathcal{H}, \mathcal{H}')$ are

$$\|A\|_1 := \mathrm{tr}\sqrt{A^*A} \qquad \text{(trace norm)} \tag{16.40}$$

$$\|A\|_2 := \sqrt{\mathrm{tr}A^*A} = \sqrt{\langle A, A\rangle_{HS}} \qquad \text{(Hilbert-Schmidt norm)} \tag{16.41}$$

We will sometimes identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ using the isomorphism

$$z \mapsto \begin{pmatrix} Re(z) \\ Im(z) \end{pmatrix} \tag{16.42}$$

For fixed Hilbert space $\mathcal{H}$, we equip the set $\mathcal{L}^+(\mathcal{H})$ with a semiorder [1] trace $\geq$. We define for each two matrices $A, B \in \mathcal{L}(\mathcal{H})$

$$A \;\geq\; B \;:\Leftrightarrow A - B \text{ positive semidefinite.} \tag{16.43}$$

The above definition also motivates to write $A \geq 0$ to indicate that A is positive semidefinite. We have

Lemma 139 (Conjugation rule)**.** *Let A, B be hermitian matrices, C a matrix. It holds*

$$A \leq B \;\Rightarrow\; CAC^* \leq CBC^*.$$

16.2 Finite-valued random variables and random matrices

The usual way, to establish events on a set Ω is to specify a *σ-Algebra*. This is a family Σ of subsets of Ω which has the following properties

- $\emptyset, \Omega \in \Sigma$,
- If $A, B \in \Sigma$, then $A \cap B \in \Sigma$, and
- If $A_1, A_2, \cdots \in \Sigma$, $\bigcup_{i=1}^{\infty} \in \Sigma$.

A *probability measure* is then a σ-additive set function, i.e. a function $\mu : \Sigma \to \mathbb{R}$ such that

[1] A semiorder is distinguished from an order by lacking the trichotomy property, i.e. not all pairs of elements can be compared

- $\mu(\emptyset) = 0$ and $\mu(\Omega) = 1$,
- For each family $\{A_i\}_{i=1}^{\infty} \subset \Sigma$, $A_i \cap A_j = \emptyset$ $i \neq j$, $\mu(\bigcap_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

The pair (Ω, Σ) is called a measureable space, the triple (Ω, Σ, μ) is then called a *probability space*. Let (Ω, Σ), (Ω', Σ') be two measurable spaces. A *random variable* is a map $X : \Omega \to \Omega'$, such that $X^{-1}(A) \in \Sigma$ for each $A \in \Sigma'$. For each set Ω, the power set forms a σ-algebra. If the set is not countable, however, the power set may be "too large to be useful". A standard σ-algebra for $\mathbb{R}$ is the so-called *Borel σ-algebra*, which is the smallest σ-algebra containing all open sets in $\mathbb{R}$. All random quantities in this course will be assumed to have only a finite range of possible values, i.e. we assume Ω to be a finite set and implicitely work with the σ-algebra formed by the members of the power set.
Therefore, we will restrain from introducing the full measure-theoretic framework for probability. Instead introducing sigma algebras and measures on the formal levels, we note, that each probability law or "probability distribution" is uniquely determined by the values on the atomic sets. We therefore define a probability distribution on a finite set $\mathcal{X}$ to be function $p : \mathcal{X} \to [0, 1]$ such that $\sum_{x \in \mathcal{X}} p(x) = 1$. The probability, that a random variable with distribution (or law) p takes a value in a set $A \subset \mathcal{X}$ is

$$\Pr(X \in A) = \sum_{x \in A} p(x). \tag{16.44}$$

The expectation of X is defined by

$$\mathbb{E}X = \sum_{x \in \mathcal{X}} p(x)x. \tag{16.45}$$

Proposition 140 (Law of large numbers)**.** *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of real random variables, $\delta > 0$. It holds*

$$\lim_{N \to \infty} \Pr\left(\left| \frac{1}{N} \sum_{i=1}^{N} X_i - \mathbb{E}X \right| \geq \delta \right) = 0. \tag{16.46}$$

16.3 Convex analysis

A set $A \subset \mathbb{R}^n$ is called *convex* , if for each elements $x_1, \ldots, x_N \in A$ and real numbers $\lambda_1, \ldots \lambda_N \in [0, 1]$ such that $\sum_{i=1}^{N} \lambda_i = 1$, the *convex combination*

$$x = \sum_{i=1}^{N} \lambda_i x_i \tag{16.47}$$

is also an element of A (i.e. "A is closed under forming finite convex combinations"). An element x of a convex set $X \in \mathbb{R}^n$ is called *extremal*, if all convex combinations representing x are trivial, i.e. if for all $b_1, b_2 \in X$, $\lambda \in [0, 1]$ either $b_1 = b_2 = x$ or $\lambda \in \{0, 1\}$ holds. Let $A \subset \mathbb{R}^d$ be a convex set. A function $f : A \to \mathbb{R}$ is called *convex* , if

$$f(\lambda a_1 + (1 - \lambda)a_2) \leq \lambda f(a_1) + (1 - \lambda) f(a_2) \tag{16.48}$$

holds for all $a_1, a_2 \in A$, $\lambda \in [0,1]$. Moreover, f is called *concave* , if $-f$ is convex, and *affine* , if it is convex and concave. We also will need the notion of a convex subset of $\mathbb{C}^n$. In this case, nothing changes, because the map

$$z \mapsto \begin{pmatrix} Re(z) \\ Im(z) \end{pmatrix} \tag{16.49}$$

is an affine bijection between $\mathbb{C}$ and $\mathbb{R}^2$.

Proposition 141 (Jensen's inequality)**.** *Let $\lambda_1, \ldots, \lambda_N \in [0,1]$, $\sum_{i=1}^N \lambda_i = 1$ and $x_1, \ldots x_N \in \mathbb{R}$. If f is convex, it holds*

$$f(\lambda_1 x_1 + \cdots + \lambda_N x_N) \leq \lambda_1 f(x_1) + \cdots + \lambda_N f(x_N). \tag{16.50}$$

In particular, if X is a real random variable and f a convex function, it holds

$$f(\mathbb{E}X) \leq \mathbb{E}f(X). \tag{16.51}$$

16.4 Exercises

Exercise 142. *Show, that the matrix product AB of two Hermitian matrices A, B is Hermitian if and only if $[A, B] = 0$.*

17 Matrix monotonicity

In this appendix, we prove matrix monotonicity of the square root function. We begin by defining matrix monotonicity Let $\mathcal{D} \subset \mathbb{R}$ and $f : \mathcal{D} \to \mathbb{R}$. Let $\hat{f}_n : \mathcal{A}_n(\mathcal{D}) \to \mathbb{H}_n$ the matrix function generated by f according to (...).

Definition 143. *$f : \mathcal{D} \to \mathbb{R}$ is* matrix monotone (matrix monotonically increasing), *if for each $n \in \mathbb{N}$ the function $\hat{f}_n$ fullfills*

$$\forall A, B \in \mathcal{A}_n(\mathcal{D}) : \; A \leq B \;\Rightarrow\; \hat{f}_n(A) \leq \hat{f}_n(B).$$

The concept of matrix monotonicity goes beyond the ordinary monotonicity of a real function, monotonicity as a real function does not imply matrix monotonicity.

Example 144. *The function $t \mapsto t^2$ is not matrix monotone on $[0, \infty]$.*

Lemma 145. *The function $t \mapsto t^{\frac{1}{2}}$ is matrix monotone.*

To prove Lemma 145, we need some supporting claims we prove first.

Index

Bibliography

[1] Igor Bjelaković and Rainer Siegmund-Schultze. Quantum steins lemma revisited, inequalities for quantum entropies, and a concavity theorem of lieb. 2012.

[2] A. Holevo. *Quantum Systems, Channels, Information - A Mathematical Introduction*. de Gruyter, Berlin, 2012.

[3] Mark M. Wilde. *Quantum Information Theory*. Camb, 2013.