Quantum information theory (MAT4430) **Spring 2021**

Lecture 9: Schumacher's compression theorem and quantum entropy

Lecturer: Alexander Müller-Hermes

Recall Shannon's source coding theorem from Lecture 1. There, we considered a discrete memoryless source of information, i.e., a sequence of independent and identically distributed random variables on a finite alphabet, and showed that it can be compressed with compression rates arbitrarily close to the Shannon entropy of the source. We will now discuss the quantum analogue of this result.

1 Compression of quantum data

Definition 1.1 (Quantum source)**.** *A* discrete memoryless source of quantum information *is a sequence* $(\rho^{\otimes n})_{n\in\mathbb{N}}$ *of tensor powers of a quantum state* $\rho \in D(\mathcal{H})$*.*

An interpretation of this concept goes as follows: Consider an ensemble $\{p_n, |\psi_n\rangle\langle\psi_n|\}_{n=1}^N$ of quantum states, which can be thought of as describing a probabilistic process that produces the pure state $|\psi_n\rangle\langle\psi_n|$ with probabilty p_n. As we have seen in Lecture 1, we can associate the quantum state $\rho = \sum_{n=1}^N p_n\, |\psi_n\rangle\langle\psi_n|$ to the outcome of this process, which takes into account that we do not know which of the states $|\psi_n\rangle\langle\psi_n|$ has been prepared. Executing this process multiple times independent from each other can then be described by the tensor powers $\rho^{\otimes n}$. Note that this is completely analogous to the classical situation, where a random variable X distributed according to a distribution $p \in \mathcal{P}(1,\dots,N)$ can be thought of as producing outcomes x_n with probability p_n.

There is, however, one major difference between the quantum case and the classical case: Entanglement! The quantum state ρ could be the reduced density matrix of a quantum system 'A' embedded in a larger quantum system 'AR', where the other quantum system 'R' serves the purpose of an environment but is often referred to as a *reference* system in this context. In this case, it will be important that the protocol compressing the system 'A' preserves the entanglem with the reference 'R'. For example, we could imagine a situation, where the researcher Alice wants to send part of an entangled quantum state to the researcher Bob while keeping her share of the state. Using a compression protocol she can make the transmission process more efficient, but of course it is important that in the end Bob can retrieve the entanglement with the system that Alice had kept. To define quantum compression schemes that preserve this entanglement we will use the so-called channel fidelity as a distance measure:

Definition 1.2 (Channel fidelity)**.** *For a quantum channel* $T : B(\mathcal{H}) \to B(\mathcal{H})$ *we define the* channel fidelity *of* T *with respect to a quantum state* $\rho \in D(\mathcal{H})$ *as*

$$F(T,\rho) = F\left((\mathrm{id}_E \otimes T)\left(|\psi_{EA}\rangle\langle\psi_{EA}|\right), |\psi_{EA}\rangle\langle\psi_{EA}|\right),$$

where $|\psi_{EA}\rangle = \mathrm{vec}\left(\sqrt{\rho}\right)$*.*

The channel fidelity $F(T,\rho)$ quantifies how close the quantum channel T is to the identity channel, when applied to part of a purification of the quantum state ρ. We will need the following elementary property of the channel fidelity:

Lemma 1.3. *Consider a quantum channel* $T : B(\mathcal{H}) \to B(\mathcal{H})$ *given by its Kraus decomposition*

$$T = \sum_{n=1}^N \mathrm{Ad}_{K_n},$$

with $K_n \in B(\mathcal{H})$. For any quantum state $\rho \in D(\mathcal{H})$ we have

$$F(T,\rho) = \sqrt{\sum_{n=1}^{N} |\langle \rho, K_n \rangle_{HS}|^2}.$$

Proof. In the exercises, we have seen that

$$F(\sigma, |\phi\rangle\langle\phi|) = \sqrt{\langle\phi|\sigma|\phi\rangle},$$

for any $\sigma \in D(\mathcal{H}')$ and any pure state $|\phi\rangle\langle\phi| \in D(\mathcal{H}')$. Applying this result, we find that

$$F(T,\rho) = \sqrt{\langle\psi_{EA}|(\mathrm{id}_E \otimes T)\left(|\psi_{EA}\rangle\langle\psi_{EA}|\right)|\psi_{EA}\rangle},$$

where $|\psi_{EA}\rangle = \mathrm{vec}\left(\sqrt{\rho}\right)$. Finally, note that

$$\langle\psi_{EA}|(\mathrm{id}_E \otimes \mathrm{Ad}_K)\left(|\psi_{EA}\rangle\langle\psi_{EA}|\right)|\psi_{EA}\rangle = |\,\mathrm{vec}(\sqrt{\rho})^\dagger \,\mathrm{vec}(K\sqrt{\rho})|^2 = |\langle\rho, K\rangle_{HS}|^2,$$

and the lemma follows from the Kraus representation

$$T = \sum_{n=1}^{N} \mathrm{Ad}_{K_n}.$$

□

Now, we define the compression task as follows:

Definition 1.4 (Quantum compression schemes)**.** *Let $\mathcal{H}$ denote a complex Euclidean space and $\rho \in D(\mathcal{H})$ a quantum state. An* (n, m, δ)-compression scheme *for ρ is a pair of quantum channels*

$$E : B(\mathcal{H}^{\otimes n}) \to B((\mathbb{C}^2)^{\otimes m}) \quad \text{and} \quad D : B((\mathbb{C}^2)^{\otimes m}) \to B(\mathcal{H}^{\otimes n}),$$

such that

$$F(D \circ E, \rho^{\otimes n}) \geqslant 1 - \delta.$$

Note that we used the channel fidelity to quantify the final error of the compression scheme, which takes entanglement with a reference system into account. Indeed, it would be trivial (and not very interesting) to construct compression schemes otherwise, since

$$\rho^{\otimes n} = (D_{\mathrm{triv}} \circ E_{\mathrm{triv}})(\rho^{\otimes n}),$$

for the quantum channels

$$E_{\mathrm{triv}} = \mathrm{Tr}\,[\cdot] \quad \text{and} \quad D_{\mathrm{triv}} = \rho^{\otimes n}.$$

Such a compression scheme would of course be useless as a compression scheme in practice, since it would destroy all correlations the quantum system might have with some other system. Note that a similar trivial example was excluded in the classical case by definition, since we only considered deterministic compression schemes. If we would have allowed for probabilistic schemes, then we would have needed to be more careful with the definition.

As in the classical case, we will also define achievable compression rates as follows:

Definition 1.5 (Achievable compression rates)**.** *We call $R \in \mathbb{R}^+$ an* achievable compression rate for $\rho \in D(\mathcal{H})$ *if for every $n \in \mathbb{N}$ there exists an (n, m_n, δ_n) compression scheme such that*

$$R = \lim_{n\to\infty} \frac{m_n}{n} \quad \text{and} \quad \lim_{n\to\infty} \delta_n = 0.$$

2 The von Neumann entropy and Schumacher compression

The von Neumann entropy is the proper quantum generalization of the Shannon entropy:

Definition 2.1 (von Neumann entropy)**.** *For a quantum state $\rho \in D(\mathcal{H})$ we define the von Neumann entropy as*

$$H(\rho) = -\operatorname{Tr}\left[\rho \log(\rho)\right],$$

where the logarithm is taken in base 2.

We will prove the following theorem:

Theorem 2.2 (Schumacher's compression theorem)**.** *Let $\mathcal{H}$ denote a complex Euclidean space and $\rho \in D(\mathcal{H})$ a quantum state.*

1. *Any number $R > H(\rho)$ is an achievable compression rate for $(\rho^{\otimes n})_{n\in\mathbb{N}}$.*
2. *If there is a sequence of (n_k, m_k, δ_k)-compression schemes for $(\rho^{\otimes n})_{n\in\mathbb{N}}$ satisfying*

$$\lim_{k\to\infty} n_k = \infty \quad \text{and} \quad \lim_{k\to\infty} \frac{m_k}{n_k} = R < H(p),$$

then we have $\lim_{k\to\infty} \delta_k = 1$, i.e., the channel fidelity of the compression schemes goes to zero in the limit $k \to \infty$.

Proof. For brevity set $d = \dim(\mathcal{H})$. We start with the direct part: By the spectral theorem, we have

$$\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i|,$$

for a probability distribution $p \in \mathcal{P}(1, \ldots, d)$ and an orthonormal basis $\{|\psi_1\rangle, \ldots, |\psi_d\rangle\} \subset \mathcal{H}$. Note that $H(\rho) = H(p)$, and recall the set of ϵ-typical strings $\mathcal{T}_{n,\epsilon}(p)$ of length n. For each $\epsilon > 0$ and $n \in \mathbb{N}$, we define an ϵ-typical projector $\Pi_{n,\epsilon} \in \operatorname{Proj}(\mathcal{H}^{\otimes n})$ by

$$\Pi_{n,\epsilon} = \sum_{(i_1,\ldots,i_n)\in\mathcal{T}_{n,\epsilon}(p)} |\psi_{i_1}\rangle\langle\psi_{i_1}| \otimes \cdots \otimes |\psi_{i_n}\rangle\langle\psi_{i_n}|.$$

Expressing the basic properties of typical sequences (see Lecture 1) in terms of the projector $\Pi_{n,\epsilon}$ shows the following:

- For any $n \in \mathbb{N}$ and any $\epsilon > 0$ we have

$$2^{-n(H(\rho)+\epsilon)}\Pi_{n,\epsilon} < \Pi_{n,\epsilon}\rho^{\otimes n}\Pi_{n,\epsilon} < 2^{-n(H(\rho)-\epsilon)}\Pi_{n,\epsilon}. \tag{1}$$

- For any $n \in \mathbb{N}$ and $\epsilon > 0$ we have

$$\operatorname{Tr}\left[\Pi_{n,\epsilon}\right] = |\mathcal{T}_{n,\epsilon}(p)| \leqslant 2^{n(H(\rho)+\epsilon)}. \tag{2}$$

- For any $\epsilon > 0$ we have

$$\lim_{n\to\infty} \operatorname{Tr}\left[\Pi_{n,\epsilon}\rho^{\otimes n}\right] = 1. \tag{3}$$

For $\epsilon > 0$ and any $n \in \mathbb{N}$, we will now construct an $(n, \lceil n(H(\rho)+\epsilon)\rceil, \delta_n)$ compression scheme for ρ such that $\delta_n \to 0$ as $n \to \infty$. This shows that $H(\rho) + \epsilon$ is an achievable rate. Consider $m = \lceil n(H(\rho)+\epsilon)\rceil$ and choose a bit string $b(i_1, i_2, \ldots, i_n) \in \{0,1\}^m$ for any typical sequence $(i_1, \ldots, i_n) \in \mathcal{T}_{\epsilon,n}(p)$. Let us denote by $\mathcal{S}_{n,m} \subset (\mathbb{C}^2)^{\otimes m}$ the span of the orthonormal vectors $|b(i_1, \ldots, i_n)\rangle$ for all $(i_1, \ldots, i_n) \in \mathcal{T}_{\epsilon,n}(p)$, and define an isometry $V_{n,\epsilon} : \mathcal{S}_{n,m} \to \mathcal{H}^{\otimes n}$ by

$$V_{n,\epsilon} = \sum_{(i_1,\ldots,i_n)\in\mathcal{T}_{n,\epsilon}(p)} \left(|\psi_{i_1}\rangle \otimes \cdots \otimes |\psi_{i_m}\rangle\right) \langle b(i_1, \ldots, i_n)|$$

It is easy to see that

$$\Pi_{n,\epsilon} = V_{n,\epsilon} V_{n,\epsilon}^\dagger.$$

Next, we define two quantum channels:

$$\begin{aligned} &E : B(\mathcal{H}^{\otimes n}) \to B((\mathbb{C}^2)^{\otimes m}) \\ &E(X) = V_{n,\epsilon}^\dagger X V_{n,\epsilon} + \operatorname{Tr}\left[(\mathbb{1}_{\mathcal{H}}^{\otimes n} - \Pi_{n,\epsilon})X\right] |0\rangle\langle 0|. \end{aligned}$$

and

$$\begin{aligned} &D : B((\mathbb{C}^2)^{\otimes m} \to B(\mathcal{H}^{\otimes n}) \\ &D(Y) = V_{n,\epsilon} Y V_{n,\epsilon}^\dagger + \operatorname{Tr}\left[(\mathbb{1}_{\mathcal{H}}^{\otimes n} - \Pi_{\mathcal{S}_{n,m}})X\right] |0\rangle\langle 0|, \end{aligned}$$

where we extend $V_{n,\epsilon}$ to all of $(\mathbb{C}^2)^{\otimes m}$ by setting it zero on the basis vectors it is not defined on. Finally, we can compute that

$$(D \circ E)(X) = \Pi_{n,\epsilon} X \Pi_{n,\epsilon} + F_{n,\epsilon}(X),$$

for some completely positive map $F_{n,\epsilon} : B(\mathcal{H}^{\otimes n}) \to B(\mathcal{H}^{\otimes n})$ that we do not need to know exactly. Using Lemma 1.3 and (3), we find that

$$F(D \circ E, \rho^{\otimes n}) \geqslant \operatorname{Tr}\left[\Pi_{n,\epsilon} \rho^{\otimes n}\right] \to 1 \text{ as } n \to \infty.$$

For the reverse direction, consider a sequence of (n_k, m_k, δ_k)-compression schemes for $\rho \in D(\mathcal{H})$ given by quantum channels

$$E_k : B(\mathcal{H}^{\otimes n_k}) \to B((\mathbb{C}^2)^{\otimes m_k}) \quad \text{and} \quad D_k : B((\mathbb{C}^2)^{\otimes m_k}) \to B(\mathcal{H}^{\otimes n_k}),$$

such that $\lim_{k\to\infty} n_k = \infty$ and

$$R = \lim_{k\to\infty} \frac{m_k}{n_k} < H(\rho).$$

Let $\{A_l^{(k)}\}_{l=1}^{L_k}$ denote the Kraus operators of the quantum channel $D_k \circ E_k$ and note that $\operatorname{rk}\left(A_l^{(k)}\right) \leqslant 2^{m_k}$ by assumption. For each k and l we denote by $\Pi_l^{(k)}$ the projection onto the image of $A_l^{(k)}$ such that $\operatorname{rk}\left(\Pi_l^{(k)}\right) \leqslant 2^{m_k}$ and $\Pi_l^{(k)} A_l^{(k)} = A_l^{(k)}$. By Lemma 1.3 we have

$$F(D_k \circ E_k, \rho^{\otimes n_k}) = \sqrt{\sum_{l=1}^{L_k} |\langle \rho^{\otimes n_k}, A_l^{(k)} \rangle_{HS}|^2}.$$

Using the Cauchy-Schwarz inequality, we can compute that

$$|\langle \rho^{\otimes n_k}, A_l^{(k)} \rangle_{HS}|^2 = |\langle \Pi_l^{(k)} \sqrt{\rho^{\otimes n_k}}, \sqrt{\rho^{\otimes n_k}} A_l^{(k)} \rangle_{HS}|^2 \leqslant \operatorname{Tr}\left[\Pi_l^{(k)} \rho^{\otimes n_k}\right] \operatorname{Tr}\left[\operatorname{Ad}_{A_l^{(n_k)}}\left(\rho^{\otimes n_k}\right)\right],$$

which implies that

$$F(D_k \circ E_k, \rho^{\otimes n_k}) \leqslant \sqrt{\sum_{l=1}^{L_k} q_l^{(k)} \operatorname{Tr}\left[\Pi_l^{(k)} \rho^{\otimes n_k}\right]}, \tag{4}$$

where we set

$$q_l^{(k)} = \operatorname{Tr}\left[\operatorname{Ad}_{A_l^{(n_k)}}\left(\rho^{\otimes n_k}\right)\right] \geqslant 0.$$

Since $D_k \circ E_k$ is a quantum channel, we find that

$$\sum_{l=1}^{L_k} q_l^{(k)} = \sum_{l=1}^{L_k} \mathrm{Tr}\left[\mathrm{Ad}_{A_l^{(n_k)}}\left(\rho^{\otimes n_k}\right)\right] = 1.$$

Finally, choosing $\epsilon > 0$ such that $R + \epsilon < H(\rho)$, we compute

$$\begin{aligned}
\mathrm{Tr}\left[\Pi_l^{(k)} \rho^{\otimes n_k}\right] &= \mathrm{Tr}\left[\Pi_l^{(k)} \Pi_{n_k,\epsilon} \rho^{\otimes n} \Pi_{n_k,\epsilon}\right] + \mathrm{Tr}\left[\Pi_l^{(k)} \rho^{\otimes n_k}\left(\mathbb{1}_{\mathcal{H}}^{\otimes n_k} - \Pi_{n_k,\epsilon}\right)\right] \\
&\leqslant 2^{-n_k(H(\rho)-\epsilon)} \,\mathrm{Tr}\left[\Pi_l^{(k)} \Pi_{n_k,\epsilon}\right] + \mathrm{Tr}\left[\rho^{\otimes n_k}\left(\mathbb{1}_{\mathcal{H}}^{\otimes n_k} - \Pi_{n_k,\epsilon}\right)\right] \\
&\leqslant 2^{n_k\left(\frac{m_k}{n_k} - H(\rho) + \epsilon\right)} + \mathrm{Tr}\left[\rho^{\otimes n_k}\left(\mathbb{1}_{\mathcal{H}}^{\otimes n_k} - \Pi_{n_k,\epsilon}\right)\right] \longrightarrow 0,
\end{aligned}$$

as $k \to \infty$, where we used that $[\rho^{\otimes n_k}, \Pi_{n_k,\epsilon}] = 0$ in the first line, (1) and $\Pi_l^{(k)} \leqslant \mathbb{1}_{\mathcal{H}}^{\otimes n_k}$ in the second line, and $\mathrm{Tr}\left[\Pi_l^{(k)} \Pi_{n_k,\epsilon}\right] \leqslant \mathrm{Tr}\left[\Pi_l^{(k)}\right] \leqslant 2^{m_k}$ and (3) in the final line. Combining this estimate with (4) shows that $F(D_k \circ E_k, \rho^{\otimes n_k}) \to 0$ as $k \to \infty$.

□

Schumacher's compression theorem gives an operational interpretation of von Neumann's entropy. We will now use a bit of time to prove some properties of the von Neumann entropy. As for the Shannon entropy, it will be useful to use the quantum relative entropy to show properties of the von Neumann entropy.

3 The quantum relative entropy

Most properties of the von Neumann entropy will follow from the properties of the quantum relative entropy and in particular from its data-processing inequality. We start by defining this quantity:

Definition 3.1 (Quantum relative entropy)**.** *Let $\mathcal{H}$ denote a complex Euclidean space. For any pair of quantum states $\rho, \sigma \in D(\mathcal{H})$ we define a the* relative entropy *as*

$$D(\rho\|\sigma) = \begin{cases} \mathrm{Tr}\left[\rho\left(\log(\rho) - \log(\sigma)\right)\right], & \text{if } \ker(\sigma) \subseteq \ker(\rho) \\ +\infty, & \text{otherwise.} \end{cases}$$

Usually, the operator $\log(\rho)$ is only defined for positive definite operators ρ. However, even if $\rho \in D(\mathcal{H})$ has a non-trivial kernel, we can define the operator $\rho\log(\rho)$ by using the convention that $0 \cdot \log(0) = 0$. Specifically, we set

$$\rho\log(\rho) = \sum_{i=1}^{n} \lambda_i \log(\lambda_i) |v_i\rangle\langle v_i|,$$

where $\rho = \sum_{i=1}^{n} \lambda_i |v_i\rangle\langle v_i|$ is the spectral decomposition of ρ with $\lambda_i > 0$ for any $i \in \{1, \ldots, n\}$ and $n = \mathrm{rk}(\rho) \leqslant \dim(\mathcal{H})$. In a similar way we can make sense of the operator $\rho\log(\sigma)$ under the condition that $\ker(\sigma) \subseteq \ker(\rho)$, by setting

$$\rho\log(\sigma) = \sum_{j=1}^{m} \log(\mu_j) \rho |w_j\rangle\langle w_j|,$$

where $\sigma = \sum_{j=1}^{m} \mu_j |w_j\rangle\langle w_j|$ is the spectral decomposition of σ with $\mu_j > 0$ for any $j \in \{1, \ldots, m\}$ and $m = \mathrm{rk}(\sigma) \leqslant \dim(\mathcal{H})$. Using the spectral decomposition of both operators, we obtain the formula

$$D(\rho\|\sigma) = \sum_{i=1}^{n} \sum_{j=1}^{m} |\langle v_i|w_j\rangle|^2 \lambda_i \left(\log(\lambda_i) - \log(\mu_j)\right), \tag{5}$$

whenever $\ker(\sigma) \subseteq \ker(\rho)$. This expression has a simple, but useful, consequence. It allows to write the relative entropy as a limit of slightly simpler trace functionals:

Lemma 3.2. *For any $\rho, \sigma \in D(\mathcal{H})$ we have*

$$D(\rho\|\sigma) = \frac{1}{\ln(2)} \lim_{\epsilon \searrow 0} \frac{1 - \mathrm{Tr}\left[\rho^{1-\epsilon}\sigma^{\epsilon}\right]}{\epsilon}.$$

Proof. If $\ker(\sigma) \not\subseteq \ker(\rho)$, then we have

$$\lim_{\epsilon \searrow 0} \left(1 - \mathrm{Tr}\left[\rho^{1-\epsilon}\sigma^{\epsilon}\right]\right) = 1 - \mathrm{Tr}\left[\rho\left(\mathbb{1}_{\mathcal{H}} - \Pi_{\ker(\sigma)}\right)\right] = \mathrm{Tr}\left[\rho\Pi_{\ker(\sigma)}\right] > 0.$$

In this case, we conclude that the limit in the statement diverges as it should.

Assume now, that $\ker(\sigma) \subseteq \ker(\rho)$ and consider the spectral decompositions

$$\rho = \sum_{i=1}^{n} \lambda_i |v_i\rangle\langle v_i|,$$

and

$$\sigma = \sum_{j=1}^{m} \mu_j |w_j\rangle\langle w_j|,$$

with $n = \mathrm{rk}(\rho)$ and $m = \mathrm{rk}(\sigma)$. Then, we define a function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(\alpha) = \mathrm{Tr}\left[\rho^{1-\alpha}\sigma^{\alpha}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} |\langle v_i|w_j\rangle|^2 \lambda_i^{1-\alpha}\mu_j^{\alpha}.$$

The function f is differentiable in every point $\alpha \in \mathbb{R}$ and it is easy to compute that

$$f'(\alpha) = -\sum_{i=1}^{n}\sum_{j=1}^{m} |\langle v_i|w_j\rangle|^2 \lambda_i^{1-\alpha}\mu_j^{\alpha}\left(\ln(\lambda_i) - \ln(\mu_j)\right).$$

Finally, we observe that $f(0) = \mathrm{Tr}[\rho] = 1$ and

$$\begin{aligned} -f'(0) &= \lim_{\epsilon \searrow 0} \frac{1 - \mathrm{Tr}\left[\rho^{1-\epsilon}\sigma^{\epsilon}\right]}{\epsilon} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} |\langle v_i|w_j\rangle|^2 \lambda_i \left(\ln(\lambda_i) - \ln(\mu_j)\right) \\ &= \ln(2) D(\rho\|\sigma). \end{aligned}$$

This finishes the proof. □

The previous lemma is quite useful when proving the data-processing inequality for the relative entropy.

4 The data-processing inequality of the relative entropy

There are at least three different ways in the literature to establish the data-processing inequality for the relative entropy. The most common one derives it from the joint convexity inequality using a technique we have seen in the exercises. Here, we will do a different approach, which directly establishes the data-processing inequality. It uses Lemma 3.2 and proves that

$$\mathrm{Tr}\left[T(\rho)^{1-\epsilon}T(\sigma)^{\epsilon}\right] \geqslant \mathrm{Tr}\left[\rho^{1-\epsilon}\sigma^{\epsilon}\right],$$

for any $\rho, \sigma \in D(\mathcal{H})$ and all quantum channels $T : B(\mathcal{H}) \to B(\mathcal{H}')$. Taking the limit as in Lemma 3.2 then yields the data-processing inequality of the relative entropy.

4.1 Monotonicity of certain Hilbert-Schmidt operators

Let $\mathcal{H}$ denote a complex Euclidean space. For positive operators $A, B \in B(\mathcal{H})^+$, we define the linear maps $L_A : B(\mathcal{H}) \to B(\mathcal{H})$ and $R_B : B(\mathcal{H}) \to B(\mathcal{H})$ by

$$L_A(X) = AX, \quad \text{and} \quad R_B(X) = XB.$$

We will now think of these maps as operators acting on the Hilbert-Schmidt inner product space $B(\mathcal{H})$. The following properties are easy to show:

- For any $A, B \in B(\mathcal{H})^+$ the operators L_A and R_B are positive[1] semidefinite, since, e.g., $L_A = L_A^*$ and $L_A = L_{\sqrt{A}} \circ L_{\sqrt{A}}$ and the same argument works for R_B.
- For any $A, B \in B(\mathcal{H})^+$ the operators L_A and R_B commute.
- For any $A, B \in B(\mathcal{H})^{++}$, the composition $R_B \circ L_A^{-1}$ is a positive semidefinite operator as well.

It is easy to simulatenously diagonalize the operators L_A and R_B by using the orthonormal basis given by $\{|v_i\rangle\langle w_j|\}_{i,j} \subset B(\mathcal{H})$ with the eigenbasis $\{|v_i\rangle\}_i \subset \mathcal{H}$ of A and the eigenbasis $\{|w_j\rangle\}_j \subset \mathcal{H}$ of B, and clearly the eigenvalues of L_A and R_B are positive.

We will now use the functional calculus for normal operators to apply functions to these operators. We start with a definition:

Definition 4.1. *For a complex Euclidean space $\mathcal{H}$ and a function $f : (0, \infty) \to (0, \infty)$, we define the operator*

$$G_f(A, B) = f\left(R_B \circ L_A^{-1}\right) \circ L_A,$$

for any pair of positive invertible operators $A, B \in B(\mathcal{H})^{++}$.

By the spectral decomposition of $R_B \circ L_A^{-1}$ in the basis $\{|v_i\rangle\langle w_j|\}_{i,j} \subset B(\mathcal{H})$ it is easy to show that the operator $G_f(A, B)$ is positive semidefinite for any $A, B \in B(\mathcal{H})^{++}$. Let us consider the operator $G_f(A, B)$ for the function $f(x) = x^\alpha$ where $\alpha \in (0, 1)$. Then, it is easy to show that

$$G_{x^\alpha}(A, B) = R_{B^\alpha} \circ L_{A^{1-\alpha}}.$$

A consequence of this is the following lemma which proof is immediate:

Lemma 4.2. *For invertible quantum states $\rho, \sigma \in D(\mathcal{H})$ we have*

$$\langle \mathbb{1}_\mathcal{H}, G_{x^\alpha}(\rho, \sigma)\mathbb{1}_\mathcal{H}\rangle_{HS} = \mathrm{Tr}\left[\rho^{1-\alpha}\sigma^\alpha\right].$$

To prove the data-processing inequality we will use the following integral representation of the function $f(x) = x^\alpha$:

$$x^\alpha = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty \frac{x}{\lambda + x} \lambda^{\alpha-1} d\lambda.$$

As the operators $G_f(A, B)$ are linear in f, we conclude that

$$\langle \mathbb{1}_\mathcal{H}, G_{x^\alpha}(\rho, \sigma)\mathbb{1}_\mathcal{H}\rangle_{HS} = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty \langle \mathbb{1}_\mathcal{H}, G_{\frac{x}{\lambda+x}}(\rho, \sigma)\mathbb{1}_\mathcal{H}\rangle_{HS} \lambda^{\alpha-1} d\lambda.$$

We will show the following theorem:

[1] Do not confuse this positivity with them being positive maps. Clearly, these operators are not positive maps since they are not Hermiticity preserving.

Theorem 4.3. *For invertible quantum states $\rho, \sigma \in D(\mathcal{H})$ and any quantum channel $T : B(\mathcal{H}) \to B(\mathcal{H}')$ for which $T(\rho)$ and $T(\sigma)$ are invertible we have*

$$G_{\frac{x}{\lambda+x}}(T(\rho), T(\sigma)) \geqslant T \circ G_{\frac{x}{\lambda+x}}(\rho, \sigma) \circ T^*,$$

for any $\lambda > 0$.

Before we prove this theorem, we will need two lemmata:

Lemma 4.4 (Yet another operator inequality)**.** *For any quantum channel $T : B(\mathcal{H}) \to B(\mathcal{H}')$, any $X \in B(\mathcal{H})$ and any invertible quantum state $\sigma \in D(\mathcal{H})$, we have*

$$T(X)T(\sigma)^{-1}T(X)^\dagger \leqslant T(X\sigma^{-1}X^\dagger).$$

Proof. Using Schur complements (see exercises) we have

$$\begin{pmatrix} \sigma & X \\ X^\dagger & X\sigma^{-1}X^\dagger \end{pmatrix} \in B(\mathcal{H} \oplus \mathcal{H})^+.$$

By complete positivity we have

$$\begin{pmatrix} T(\sigma) & T(X) \\ T(X)^\dagger & T(X\sigma^{-1}X^\dagger) \end{pmatrix} \in B(\mathcal{H}' \oplus \mathcal{H}')^+,$$

which, by taking Schur complements again, is equivalent to the desired operator inequality. □

Lemma 4.5. *Let $\mathcal{H}$ denote a complex Euclidean space. For positive invertible operators $A, B \in B(\mathcal{H})^{++}$ and $X \in B(\mathcal{H})$ we have*

$$XA^{-1}X^\dagger \leqslant B^{-1} \quad \text{if and only if} \quad X^\dagger BX \leqslant A.$$

Proof. Obviously, it is enough to show one direction of the equivalence. If $XA^{-1}X^\dagger \leqslant B^{-1}$, then we have $B^{1/2}XA^{-1}X^\dagger B^{1/2} \leqslant \mathbb{1}_{\mathcal{H}}$. This final condition is equivalent to $\|A^{-1/2}X^\dagger B^{1/2}\|_\infty \leqslant 1$, which implies $(A^{-1/2}X^\dagger B^{1/2})(B^{1/2}XA^{-1/2}) \leqslant \mathbb{1}_{\mathcal{H}}$. Multiplying by $A^{1/2}$ on both sides, shows that $X^\dagger BX \leqslant A$. □

Proof of Theorem 4.3. Fix $\lambda > 0$. By Lemma 4.5 it is sufficient to show that

$$G_{\frac{x}{\lambda+x}}(\rho, \sigma)^{-1} \geqslant T^* \circ G_{\frac{x}{\lambda+x}}(T(\rho), T(\sigma))^{-1} \circ T.$$

Note that

$$G_{\frac{x}{\lambda+x}}(\rho, \sigma)^{-1} = (\lambda + R_\sigma \circ L_\rho^{-1}) \circ R_\sigma^{-1} \circ L_\rho \circ L_\rho^{-1} = \lambda R_\sigma^{-1} + L_\rho^{-1},$$

and a similar expression holds for the operator $G_{\frac{x}{\lambda+x}}(T(\rho), T(\sigma))^{-1}$. With this we have

$$\begin{aligned} \langle X, G_{\frac{x}{\lambda+x}}(\rho, \sigma)^{-1}(X)\rangle_{HS} &= \operatorname{Tr}\left[X^\dagger \left(\lambda R_\sigma^{-1} + L_\rho^{-1}\right)(X)\right] \\ &= \lambda \operatorname{Tr}\left[X\sigma^{-1}X^\dagger\right] + \operatorname{Tr}\left[X^\dagger\rho^{-1}X\right] \end{aligned}$$

and

$$\begin{aligned} \langle X, T^* \circ G_{\frac{x}{\lambda+x}}(T(\rho), T(\sigma))^{-1} \circ T(X)\rangle_{HS} &= \operatorname{Tr}\left[T(X)^\dagger \left(\lambda R_{T(\sigma)}^{-1} + L_{T(\rho)}^{-1}\right)(T(X))\right] \\ = \lambda \operatorname{Tr}\left[T(X)T(\sigma)^{-1}T(X)^\dagger\right] &+ \operatorname{Tr}\left[T(X)^\dagger T(\rho)^{-1}T(X)\right], \end{aligned}$$

By Lemma 4.4, we have

$$\operatorname{Tr}\left[T(X)T(\sigma)^{-1}T(X)^\dagger\right] \leqslant \operatorname{Tr}\left[X\sigma^{-1}X^\dagger\right],$$

and

$$\operatorname{Tr}\left[T(X)^\dagger T(\rho)^{-1}T(X)\right] \leqslant \operatorname{Tr}\left[X^\dagger\rho^{-1}X\right],$$

for all invertible $\rho, \sigma \in D(\mathcal{H})$ and all $X \in B(\mathcal{H})$. □

4.2 Proving the data-processing inequality

Now, we are ready to prove the main result of this lecture:

Theorem 4.6 (Data-processing inequality)**.** *For any quantum channel $T : B(\mathcal{H}) \to B(\mathcal{H}')$, we have*

$$D(T(\rho), T(\sigma)) \leqslant D(\rho, \sigma),$$

for all quantum states $\rho, \sigma \in D(\mathcal{H})$.

Proof. Assume first that $\rho, \sigma \in D(\mathcal{H})$ and $T(\rho), T(\sigma) \in D(\mathcal{H}')$ are invertible. Then, we have

$$\operatorname{Tr}\left[\rho^{1-\alpha}\sigma^{\alpha}\right] = \langle \mathbb{1}_{\mathcal{H}}, G_{x^{\alpha}}(\rho,\sigma)\mathbb{1}_{\mathcal{H}}\rangle_{HS} = \frac{\sin(\pi\alpha)}{\pi}\int_0^{\infty} \langle \mathbb{1}_{\mathcal{H}}, G_{\frac{x}{\lambda+x}}(\rho,\sigma)\mathbb{1}_{\mathcal{H}}\rangle_{HS}\lambda^{\alpha-1}d\lambda,$$

for any $\alpha \in (0,1)$. By Theorem 4.3 we have

$$\begin{aligned}\langle \mathbb{1}_{\mathcal{H}'}, G_{\frac{x}{\lambda+x}}(T(\rho),T(\sigma))\mathbb{1}_{\mathcal{H}'}\rangle_{HS} &= \langle \mathbb{1}_{\mathcal{H}}, T\circ G_{\frac{x}{\lambda+x}}(T(\rho),T(\sigma))\circ T^{*}(\mathbb{1}_{\mathcal{H}})\rangle_{HS} \\ &\geqslant \langle \mathbb{1}_{\mathcal{H}}, G_{\frac{x}{\lambda+x}}(\rho,\sigma)\mathbb{1}_{\mathcal{H}}\rangle_{HS},\end{aligned}$$

and, using that $\sin(\pi\alpha) \geqslant 0$ for all $\alpha \in (0,1)$, we conclude that

$$\operatorname{Tr}\left[T(\rho)^{1-\alpha}T(\sigma)^{\alpha}\right] \geqslant \operatorname{Tr}\left[\rho^{1-\alpha}\sigma^{\alpha}\right],$$

for any $\alpha \in (0,1)$.

Next, consider general $\rho, \sigma \in D(\mathcal{H})$ and a quantum channel $T : B(\mathcal{H}) \to B(\mathcal{H}')$. Consider quantum states

$$\rho_{\epsilon} = (1-\epsilon)\rho + \epsilon\frac{\mathbb{1}_{\mathcal{H}}}{\dim(\mathcal{H})} \quad \text{and} \quad \sigma_{\epsilon} = (1-\epsilon)\sigma + \epsilon\frac{\mathbb{1}_{\mathcal{H}}}{\dim(\mathcal{H})},$$

for any $\epsilon > 0$, and quantum channels

$$T_{\delta} = (1-\delta)T + \delta \operatorname{Tr}[\cdot]\frac{\mathbb{1}_{\mathcal{H}}}{\dim(\mathcal{H})},$$

for any $\delta > 0$. By the computation from above, we have

$$\operatorname{Tr}\left[T_{\delta}(\rho_{\epsilon})^{1-\alpha}T_{\delta}(\sigma_{\epsilon})^{\alpha}\right] \geqslant \operatorname{Tr}\left[\rho_{\epsilon}^{1-\alpha}\sigma_{\epsilon}^{\alpha}\right],$$

for any $\alpha \in (0,1)$ and any $\epsilon, \delta > 0$. Using that the function $(\rho,\sigma) \mapsto \operatorname{Tr}\left[\rho^{1-\alpha}\sigma^{\alpha}\right]$ is continuous for every $\alpha \in (0,1)$, we conclude that

$$\operatorname{Tr}\left[T(\rho)^{1-\alpha}T(\sigma)^{\alpha}\right] = \lim_{\epsilon,\delta\searrow 0}\operatorname{Tr}\left[T_{\delta}(\rho_{\epsilon})^{1-\alpha}T_{\delta}(\sigma_{\epsilon})^{\alpha}\right] \geqslant \lim_{\epsilon\searrow 0}\operatorname{Tr}\left[\rho_{\epsilon}^{1-\alpha}\sigma_{\epsilon}^{\alpha}\right] = \operatorname{Tr}\left[\rho^{1-\alpha}\sigma^{\alpha}\right].$$

Finally, we can use Lemma 3.2 and conclude that

$$\begin{aligned}D(T(\rho)\|T(\sigma)) &= \frac{1}{\ln(2)}\lim_{\epsilon\searrow 0}\frac{1-\operatorname{Tr}\left[T(\rho)^{1-\epsilon}T(\sigma)^{\epsilon}\right]}{\epsilon} \\ &\leqslant \frac{1}{\ln(2)}\lim_{\epsilon\searrow 0}\frac{1-\operatorname{Tr}\left[\rho^{1-\epsilon}\sigma^{\epsilon}\right]}{\epsilon} \\ &= D(\rho\|\sigma).\end{aligned}$$

□

4.3 Generalizing the data-processing inequality

If you carefully read the proof given in the previous sections, you might realize that we did not use that the linear map $T : B(\mathcal{H}) \to B(\mathcal{H}')$ was a quantum channel. In the way, we have stated it, we have only used that the map $T : B(\mathcal{H}) \to B(\mathcal{H}')$ is a trace-preserving 2-positive map. By a slight modification of the proof, we can actually make it work for duals of so-called *unital Schwarz maps*.

Definition 4.7. *A linear map $P : B(\mathcal{H}) \to B(\mathcal{H}')$ is called a* unital Schwarz map *if it is unital and satisfies the Schwarz inequality*

$$P(X)^\dagger P(X) \leqslant P(X^\dagger X),$$

for any $X \in B(\mathcal{H})$.

Note that the Schwarz inequality for a unital map $P : B(\mathcal{H}) \to B(\mathcal{H}')$ is equivalent to

$$\begin{pmatrix} \mathbb{1}_{\mathcal{H}'} & P(X) \\ P(X)^\dagger & P(X^\dagger X) \end{pmatrix} \geqslant 0,$$

for any $X \in B(\mathcal{H})$. The following lemma is therefore immediate:

Lemma 4.8. *If $T : B(\mathcal{H}') \to B(\mathcal{H})$ is a quantum channel, then T^* is a unital Schwarz map.*

There are many examples of unital Schwarz maps that do not arise as adjoints of quantum channels, or even of 2-positive maps. The most prominent example is the map $P : B(\mathbb{C}^2) \to B(\mathbb{C}^2)$ given by

$$P(X) = \frac{1}{2}\operatorname{Tr}[X]\frac{\mathbb{1}_{\mathbb{C}^2}}{2} + \frac{1}{2}X^T.$$

It turns out, that Theorem 4.3 also holds for linear maps T that are adjoints of unital Schwarz maps. The proof is almost the same, but in the final lines of the proof we use the following inequality:

Theorem 4.9 (A tracial inequality)**.** *Let $P : B(\mathcal{H}) \to B(\mathcal{H}')$ denote a unital Schwarz map. For any $C \in B(\mathcal{H}')^+$ and any $X \in B(\mathcal{H}')$ satisfying $\ker(C) \subseteq \ker(X^\dagger)$ we have*

$$\operatorname{Tr}\left[P^*(X)^\dagger P^*(C)^{-1} P^*(X)\right] \leqslant \operatorname{Tr}\left[X^\dagger C^{-1} X\right],$$

where we used the Moore-Penrose pseudoinverse.

Proof. Setting $A = P^*(C)^{-1}P^*(X)$ with the Moore-Penrose pseudoinverse, we find that

$$\begin{pmatrix} AA^\dagger & -A \\ -A^\dagger & \mathbb{1}_{\mathcal{H}} \end{pmatrix} \begin{pmatrix} P^*(C) & P^*(X) \\ P^*(X)^\dagger & P^*\left(X^\dagger C^{-1} X\right) \end{pmatrix} = \begin{pmatrix} Z & \star \\ \star & D \end{pmatrix},$$

with

$$D = P^*(X^\dagger C^{-1} X) - P^*(X)^\dagger P^*(C)^{-1} P^*(X),$$

and

$$Z = P^*(C)^{-1}P^*(X)P^*(X)^\dagger P^*(C)^{-1}P^*(C) - P^*(C)^{-1}P^*(X)P^*(X)^\dagger.$$

Using that $P^*(C)^{-1}P^*(C)P^*(C)^{-1} = P^*(C)$, we find that $\operatorname{Tr}[Z] = 0$, and we conclude

$$\operatorname{Tr}\left[\begin{pmatrix} AA^\dagger & -A \\ -A^\dagger & \mathbb{1}_{\mathcal{H}} \end{pmatrix} \begin{pmatrix} P^*(C) & P^*(X) \\ P^*(X)^\dagger & P^*\left(X^\dagger C^{-1} X\right) \end{pmatrix}\right] = \operatorname{Tr}\left[P^*(X^\dagger C^{-1} X) - P^*(X)^\dagger P^*(C)^{-1} P^*(X)\right],$$

which is non-negative since

$$\operatorname{Tr}\left[\begin{pmatrix} AA^\dagger & -A \\ -A^\dagger & \mathbb{1}_{\mathcal{H}} \end{pmatrix} \begin{pmatrix} P^*(C) & P^*(X) \\ P^*(X)^\dagger & P^*\left(X^\dagger C^{-1} X\right) \end{pmatrix}\right] = \operatorname{Tr}\left[\begin{pmatrix} P(AA^\dagger) & -P(A) \\ -P(A)^\dagger & \mathbb{1}_{\mathcal{H}'} \end{pmatrix} \begin{pmatrix} C & X \\ X^\dagger & X^\dagger C^{-1} X \end{pmatrix}\right],$$

is the Hilbert-Schmidt inner product of two positive semidefinite operators. Finally, note that P^* is trace-preserving, and we obtain the desired inequality.

□

The data-processing inequality for the quantum relative entropy therefore holds for all trace-preserving maps that are adjoints of unital Schwarz maps. Recently, a different technique for proving the data-processing inequality was discovered. Unfortunately, it goes beyond the scope of this course, but for completeness we state the most general form of the data-processing inequality:

Theorem 4.10 (General data-processing inequality)**.** *For any positive and trace-preserving map $P : B(\mathcal{H}) \to B(\mathcal{H}')$, we have*

$$D(P(\rho), P(\sigma)) \leqslant D(\rho, \sigma),$$

for all quantum states $\rho, \sigma \in D(\mathcal{H})$.

Whether this theorem can be proven using the techniques used above is an open question!