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

Lecture 10: Encoding classical information into quantum states

Lecturer: Alexander Müller-Hermes

In this lecture, we will study how classical information can be encoded into quantum states. Specifically, we will consider a d-dimensional quantum system and define a classical channel by first preparing it in some quantum state associated to a classical message, and then measuring the system with some POVM. We will then prove Holevo's theorem giving an upper bound for the classical capacity of any such channel. It will turn out that this capacity is always less or equal to $\log(d)$, and we conclude that a d-dimensional quantum system cannot store more than $\log(d)$ classical bits of information reliably.

1 Pinsker's inequality

The following theorem will be proved in the exercises:

Theorem 1.1 (Pinsker's inequality)**.** *For any quantum states $\rho, \sigma \in D(\mathcal{H})$ we have*

$$D(\rho\|\sigma) \geqslant \frac{1}{2\ln(2)}\|\rho-\sigma\|_1^2.$$

Pinsker's inequality can be seen as a refinement of Klein's inequality (see exercises), and it shows that the relative entropy is faithful, i.e., $D(\rho\|\sigma)=0$ holds if and only if $\rho=\sigma$.

2 The quantum mutual information

The quantum mutual information generalizes the classical mutual information:

Definition 2.1 (Quantum mutual information)**.** *For a quantum state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ we define*

$$I(A:B)_{\rho_{AB}} = H(\rho_A) + H(\rho_B) - H(\rho_{AB}).$$

Like its classical analogon, we can express the quantum mutual information in terms of the relative entropy:

Lemma 2.2. *For a quantum state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ we have*

$$I(A:B)_{\rho_{AB}} = D(\rho_{AB}\|\rho_A \otimes \rho_B).$$

Proof. Exercise. □

Consequences of the previous lemma are the following elementary properties of the quantum mutual information: For any $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ we have

- $I(A:B)_{\rho_{AB}} \geqslant 0$ with equality if and only if $\rho_{AB} = \rho_A \otimes \rho_B$.
- $I(A:B)_{(\mathrm{id}_A \otimes T)(\rho_{AB})} \leqslant I(A:B)_{\rho_{AB}}$ for any quantum channel $T: B(\mathcal{H}_A) \to B(\mathcal{H}_C)$.

We can reformulate the strong subadditivity inequality in the following way:

Theorem 2.3. *For any quantum state $\rho_{ABC} \in D(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ we have*

$$I(A:B)_{\rho_{AB}} \leqslant I(A:(B,C))_{\rho_{AB}}.$$

Proof. Exercises. □

3 Convex structure of quantum measurements

Let $\mathcal{H}$ denote a complex Euclidean space and consider the set of all POVMs with at most $N \in \mathbb{N}$ outcomes, which we denote by

$$\mathcal{M}_N = \{\mu : \{1, \ldots, N\} \to B(\mathcal{H})^+ \ : \ \sum_{n=1}^{N} \mu(n) = \mathbb{1}_{\mathcal{H}}\}.$$

We will need a few basic properties of this set:

Lemma 3.1. *The set $\mathcal{M}_N$ is a compact and convex subset of $B(\mathcal{H})^N$.*

Proof. Clearly, $\mathcal{M}_N$ is convex. To show that $\mathcal{M}_N$ is compact, note that every $\mu \in \mathcal{M}_N$ gives rise to a quantum state

$$\rho_\mu = \frac{1}{N} \begin{pmatrix} \mu(1) & & & & \\ & \mu(2) & & & \\ & & \mu(3) & & \\ & & & \ddots & \\ & & & & \mu(N) \end{pmatrix} \in D(\mathbb{C}^N \otimes \mathcal{H}).$$

Since the set $D(\mathbb{C}^N \otimes \mathcal{H})$ is compact and the restriction to the diagonal blocks is continuous, we conclude that $\mathcal{M}_N$ is compact as the image of a compact set under a continuous map. □

Lemma 3.2. *Let $\mathcal{H}$ denote a complex Euclidean space and $\mu : \{1, \ldots, N\} \to B(\mathcal{H})^+$ a POVM. If $\mu \in \mathrm{Ext}\,(\mathcal{M}_N)$ is extremal, then we have*

$$|\{n \in \{1, \ldots, N\} \ : \ \mu(n) \neq 0\}| \leqslant \dim(\mathcal{H})^2.$$

Proof. Assume that

$$|\{n \in \{1, \ldots, N\} \ : \ \mu(n) \neq 0\}| > \dim(\mathcal{H})^2.$$

Then, the operators $\{\mu(1), \ldots, \mu(N)\}$ are linearly dependent and there exist $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$, not all of which are zero, such that

$$\sum_{n=1}^{N} \alpha_n \mu(n) = 0.$$

Next, we set $\alpha_{\max} = \max\{|\alpha_n| \ : \ n \in \{1, \ldots, N\}\}$ and define $\mu_0, \mu_1 : \{1, \ldots, N\} \to B(\mathcal{H})$ by

$$\mu_0(n) = \mu(n) + \frac{1}{\alpha_{\max}} \alpha_n \mu(n)$$

and

$$\mu_1(n) = \mu(n) - \frac{1}{\alpha_{\max}} \alpha_n \mu(n).$$

Note that $\mu_0(n), \mu_1(n) \in B(\mathcal{H})^+$ for every $n \in \{1, \ldots, N\}$ and that

$$\sum_{n=1}^{N} \mu_0(n) = \sum_{n=1}^{N} \mu_1(n) = \mathbb{1}_{\mathcal{H}} \pm \frac{1}{\alpha_{\max}} \sum_{n=1}^{N} \alpha_n \mu(n) = \mathbb{1}_{\mathcal{H}}.$$

With this we conclude that $\mu_0, \mu_1 \in \mathcal{M}_N$. Since

$$\mu = \frac{1}{2}(\mu_0 + \mu_1),$$

we find that μ is not an extreme point of $\mathcal{M}_N$.

□

4 The setting of Holevo's theorem

To explain the setting of Holevo's theorem, we consider a set of quantum states

$$\{\rho_x \ : \ x \in \Sigma_A\} \subset D(\mathcal{H}),$$

labeled by some alphabet Σ_A. For any alphabet Σ_B and any POVM $\mu : \Sigma_B \to B(\mathcal{H})^+$, we may define a classical communication channel $N_\mu : \Sigma_A \to \mathcal{P}(\Sigma_B)$ by

$$N_\mu(y|x) = \langle \mu(y), \rho_x \rangle_{HS}.$$

Note that we use the notation $N(y|x)$ for a classical channel $N : \Sigma_A \to \mathcal{P}(\Sigma_B)$ to denote the probability of $y \in \Sigma_B$ with respect to the probability distribution $N(x)$ (as in Lecture 1). Recall the mutual information of a joint probability distribution $p_{AB} \in \mathcal{P}(\Sigma_A \times \Sigma_B)$ given by

$$I(A:B)_{p_{AB}} = H(p_A) + H(p_B) - H(p_{AB}),$$

where p_A and p_B denote the marginals of p_{AB}. In Lecture 1, we stated Shannon's channel coding theorem:

Theorem 4.1 (Shannon's channel coding theorem)**.** *For alphabets Σ_A and Σ_B let $N : \Sigma_A \to \mathcal{P}(\Sigma_B)$ denote a communication channel. The capacity of N is given by*

$$C(N) = \sup_{p_A \in \mathcal{P}(\Sigma_A)} I(A:B)_{p_{AB}^N},$$

where

$$p_{AB}^N(x,y) = p_A(x)N(y|x),$$

is a joint probability distribution on $\Sigma_A \times \Sigma_B$ defined for any probability distribution $p_A \in \mathcal{P}(\Sigma_A)$. A rate R is achievable for communication via N if and only if

$$R < C(N).$$

Motivated by this theorem, we define the following quantity:

Definition 4.2 (Accessible information)**.** *The* accessible information *of an ensemble $\{p(x), \rho_x\}_{x \in \Sigma_A}$, where $p \in \mathcal{P}(\Sigma_A)$ and $\rho_x \in D(\mathcal{H})$ for all $x \in \Sigma_A$ is given by*

$$I_{acc}(\{p(x), \rho_x\}) = \sup_\mu I(A:B)_{p_{AB}^{N_\mu}},$$

where the supremum goes over all POVMs $\mu : \Sigma_B \to B(\mathcal{H})^+$ and all alphabets Σ_B.

The accessible information quantifies the highes possible mutual information between the input variable $x \in \Sigma_A$ and the measurement outcome $y \in \Sigma_B$ for any choice of the measurement. For any given set of quantum states $\{\rho_x \ : \ x \in \Sigma_A\} \subset D(\mathcal{H})$ we conclude that

$$\sup_\mu C(N_\mu) = \sup_{p_A \in \mathcal{P}(\Sigma_A)} I_{\text{acc}}(\{p(x), \rho_x\}).$$

This quantity equals the supremum of achievable rates for information transmission through the preparation-measurement process for an optimal choice of measurement operators. In the following, we will analyze the accessible information in more detail:

5 Accessible information is attained

We will start by showing that the supremum in the definition of the accessible information is always attained by some measurement with a finite number of measurement outcomes. To show this, we will start by showing that the optimization in the accessible information I_{acc} is over a convex function.

Lemma 5.1. *Let $\{p(x), \rho_x\}_{x\in\Sigma_A}$ denote an ensemble for $p \in \mathcal{P}(\Sigma_A)$ and quantum states $\rho_x \in D(\mathcal{H})$. For some finite alphabet Σ_B, any pair of POVMs $\mu_0, \mu_1 : \Sigma_B \to B(\mathcal{H})^+$, and any $\lambda \in [0,1]$ we have*

$$I(A:B)_{p_{AB}^{N_\mu}} \leqslant (1-\lambda) I(A:B)_{p_{AB}^{N_{\mu_0}}} + \lambda I(A:B)_{p_{AB}^{N_{\mu_1}}},$$

where

$$\mu = (1-\lambda)\mu_0 + \lambda\mu_1.$$

Proof. Recall from the exercises that

$$I(A:B)_{p_{AB}} = D\left(p_{AB} \| p_A \times p_B\right),$$

and that the function $N \mapsto I(A:B)_{p_{AB}^N}$ is convex (by joint convexity of the classical relative entropy). The statement of the lemma is an immediate consequence of this fact. □

Now, we can show that the supremum in the definition of I_{acc} is always attained by an extremal POVM.

Theorem 5.2. *Let $\{p(x), \rho_x\}_{x\in\Sigma_A}$ denote an ensemble for $p \in \mathcal{P}(\Sigma_A)$ and quantum states $\rho_x \in D(\mathcal{H})$. There exists an alphabet Σ_B with $|\Sigma_B| \leqslant \dim(\mathcal{H})^2$ and a POVM $\mu : \Sigma_B \to B(\mathcal{H})^+$ such that*

$$I_{acc}\left(\{p(x), \rho_x\}\right) = I(A:B)_{p_{AB}^{N_\mu}}.$$

Proof. Recall the set $\mathcal{M}_L \subset B(\mathcal{H})^L$ of POVMs with at most L outcomes. It will be convenient to introduce the quantities

$$I_{\text{acc}}\left(\{p(x), \rho_x\}, L\right) = \sup_{\mu\in\mathcal{M}_L} I(A:B)_{p_{AB}^{N_\mu}}, \tag{1}$$

for every $L \in \mathbb{N}$. Since the value of $I(A:B)_{p_{AB}^{N_\mu}}$ for any $\mu : \Sigma_B \to B(\mathcal{H})^+$ only depends on the joint probability distribution $p_{AB}^{N_\mu}$ and not on any details of the alphabet Σ_B, we may identify any alphabet Σ_B with the alphabet $\{1, \ldots, |\Sigma_B|\}$. This shows that

$$I_{\text{acc}}\left(\{p(x), \rho_x\}\right) = \sup_{L\in\mathbb{N}} I_{\text{acc}}\left(\{p(x), \rho_x\}, L\right).$$

For any $L \in \mathbb{N}$ the set $\mathcal{M}_L$ of POVMs $\mu : \{1, \ldots, L\} \to B(\mathcal{H})^+$ is compact and convex by Lemma 3.1. We conclude that for any $L \in \mathbb{N}$ there exists an extreme point $\mu_{opt} \in \mathcal{M}_L$ attaining the supremum in (1), i.e., such that

$$I_{\text{acc}}\left(\{p(x), \rho_x\}, L\right) = I(A:B)_{p_{AB}^{N_{\mu_{opt}}}}.$$

By Lemma 3.2 we have that

$$|\{y \in \{1, \ldots, L\} \ : \ \mu_{opt}(y) \neq 0\}| \leqslant \dim\left(\mathcal{H}\right)^2.$$

By renaming the elements $y \in \{1, \ldots, L\}$ for which $\mu_{opt}(y) \neq 0$, we can identify μ_{opt} with a POVM in $\mathcal{M}_{\dim(\mathcal{H})^2}$. This shows that

$$I(A:B)_{p_{AB}^{N_{\mu_{opt}}}} \leqslant I_{\text{acc}}\left(\{p(x), \rho_x\}, \dim\left(\mathcal{H}\right)^2\right).$$

Combining the previous statements shows that

$$I_{\mathrm{acc}}\left(\{p(x),\rho_x\}\right) = I_{\mathrm{acc}}\left(\{p(x),\rho_x\},\dim\left(\mathcal{H}\right)^2\right) = I(A:B)_{p_{AB}^{N_{\mu_{opt}}}},$$

for some $\mu_{opt} \in \mathcal{M}_{\dim(\mathcal{H})^2}$. □

6 Holevo's quantity and theorem

When studying how classical information can be encoded into quantum systems, it is very useful to consider so-called *classical-quantum states*. A quantum state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ is a classical-quantum state, if it is of the form

$$\rho_{AB}^{(cq)} = \sum_{x\in\Sigma_A} p(x)|x\rangle\langle x|_A \otimes \sigma_x^B,$$

for some probability distribution $p \in \mathcal{P}(\Sigma_A)$, quantum states $\sigma_x \in D(\mathcal{H}_B)$, and where $\{|x\rangle\}_{x\in\Sigma_A}$ denotes the computational basis. The physical interpretation of the state $\rho_{AB}^{(cq)}$ is that system 'A' is in a classical state x with probability $p(x)$. The classical state x is represented by the pure state $|x\rangle\langle x|_A$ in the computational basis. Formally, the system 'A' is still a quantum system, but since the classical information is represented in a fixed basis it can be accessed deterministically (by the PVM $\{|x\rangle\langle x|\}_{x\in\Sigma_A}$).

To any ensemble $\{p(x),\rho_x\}_{x\in\Sigma_A}$ with $p \in \mathcal{P}(\Sigma_A)$ and quantum states $\rho_x \in D(\mathcal{H})$ we may associate the classical-quantum state

$$\rho_{CA}^{(cq)} = \sum_{x\in\Sigma_A} p(x)|x\rangle\langle x|_C \otimes \rho_x^A.$$

It is straightforward to compute the mutual information

$$I(C:A)_{\rho_{CA}^{cq}} = H\left(\sum_{x\in\Sigma_A} p_x\rho_x\right) - \sum_{x\in\Sigma_B} p_x H(\rho_x),$$

which is non-negative by concavity of the von Neumann entropy. Intuitively, one might expect that this mutual information (which generalizes the classical mutual information) could somehow quantify the information that the quantum system 'A' has about the classical state x. This intuition is indeed correct, and as a result, the above quantity got its own name:

Definition 6.1 (Holevo information)**.** *For any ensemble $\{p(x),\rho_x\}_{x\in\Sigma_A}$ with $p \in \mathcal{P}(\Sigma_A)$ and quantum states $\rho_x \in D(\mathcal{H})$, we define*

$$\chi\left(\{p(x),\rho_x\}\right) = H\left(\sum_{x\in\Sigma_A} p_x\rho_x\right) - \sum_{x\in\Sigma_A} p_x H(\rho_x).$$

We will now show the following theorem:

Theorem 6.2 (Holevo's theorem)**.** *For any ensemble $\{p(x),\rho_x\}_{x\in\Sigma_A}$ with $p \in \mathcal{P}(\Sigma_A)$ and quantum states $\rho_x \in D(\mathcal{H})$, we have*

$$I_{acc}\left(\{p(x),\rho_x\}\right) \leqslant \chi\left(\{p(x),\rho_x\}\right).$$

Proof. Define the classical-quantum state

$$\sigma_{CA} = \sum_{x\in\Sigma_A} p(x)|x\rangle\langle x|_C \otimes \rho_x^A,$$

such that

$$\chi\left(\{p(x),\rho_x\}\right) = I(C:A)_{\sigma_{CA}} = D(\sigma_{CA}\|\sigma_C\otimes\sigma_A).$$

Next, consider a POVM $\mu : \{1,\ldots,L\} \to B(\mathcal{H})^+$ and define a quantum channel $M_\mu : B(\mathcal{H}) \to B(\mathbb{C}^L)$ by

$$M_\mu(X) = \sum_{y=1}^{L} \langle \mu(y), X\rangle_{HS} |y\rangle\langle y|,$$

and note that

$$\left(\mathrm{id}_A \otimes M_\mu\right)(\sigma_{CA}) = \sum_{x\in\Sigma_A}\sum_{y\in\Sigma_B} p(x)N_\mu(y|x)|x\rangle\langle x|\otimes|y\rangle\langle y| = \sum_{x\in\Sigma_A}\sum_{y\in\Sigma_B} p_{AB}^{N_\mu}(x,y)|x\rangle\langle x|\otimes|y\rangle\langle y|.$$

Similarly, we note that

$$M_\mu\left(\sigma_A\right) = \sum_{x\in\Sigma_A}\sum_{y\in\Sigma_B} p(x)N_\mu(y|x)|y\rangle\langle y| = \sum_{y\in\Sigma_B} p_B^{N_\mu}(y)|y\rangle\langle y|,$$

and, by the data-processing inequality, we conclude that

$$\begin{aligned} I(A:B)_{p_{AB}^{N_\mu}} &= D\left(\left(\mathrm{id}_A\otimes M_\mu\right)(\sigma_{CA})\,\|\sigma_C\otimes M_\mu\left(\sigma_A\right)\right) \\ &\leqslant D\left(\sigma_{CA}\|\sigma_C\otimes\sigma_A\right) = \chi\left(\{p(x),\rho_x\}\right). \end{aligned}$$

Since the POVM μ was arbitrary, the proof is finished. □

Holevo's theorem has a simple but important corollary:

Corollary 6.3. *For any ensemble $\{p(x),\rho_x\}_{x\in\Sigma_A}$ with $p\in\mathcal{P}(\Sigma_A)$ and quantum states $\rho_x\in D(\mathcal{H})$, we have*

$$I_{acc}\left(\{p(x),\rho_x\}\right) \leqslant \log(\dim(\mathcal{H})).$$

Proof. This follows by noting that

$$I_{\mathrm{acc}}\left(\{p(x),\rho_x\}\right) \leqslant \chi\left(\{p(x),\rho_x\}\right) = H\left(\sum_{x\in\Sigma_A} p_x\rho_x\right) - \sum_{x\in\Sigma_B} p_x H(\rho_x) \leqslant \log\left(\dim(\mathcal{H})\right).$$

□

Although any quantum system can be in a continuum of different quantum states, the previous corollary shows that classical messages can only be send reliably at rates below $\log\left(\dim(\mathcal{H})\right)$ by the process of preparation and measurement. This can be interpreted as saying that a qubit can only store 1 bit of information reliably.