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

Lecture 12: The Holevo-Schumacher-Westmoreland theorem

Lecturer: Alexander Müller-Hermes

In the previous lecture, we have introduced the classical capacity of quantum channels, which quantifies the ability of such a channel to transmit classical information. We have seen that the bound from Holevo's theorem can be attained by taking product states as input to many copies of the channel but allowing for global POVMs to be applied to many communication lines. In particular, we have seen that such schemes can achieve the Holevo information

$$\chi(T) = \sup_{\{p(x),\rho_x\}_{x\in\Sigma}} \chi\left(\{p(x), T(\rho_x)\}_{x\in\Sigma}\right).$$

of a quantum channel $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$. Building on this result, we will now prove the main theorem about the classical capacity of quantum channels, which is known as the Holevo-Schumacher-Westmoreland theorem due to its discoverers. After proving this theorem, we will discuss the problem of computing the classical capacity, and we will identify classes of quantum channels for which it can be evaluated.

1 The Holevo-Schumacher-Westmoreland theorem

We will need the following lemma, which we will prove in the exercises:

Lemma 1.1. *For $k \in \{1,\dots,N\}$ and $\delta \leqslant \frac{1}{k}$ let $\mathcal{P}_{\delta,k}(N)$ denote the set of all probability distributions $p \in \mathcal{P}(\{1,\dots,N\})$ such that $p(1),\cdots,p(k) \geqslant \delta$. Then, we have*

$$\max_{p\in\mathcal{P}_\delta(N)} = -k\delta\log(\delta) - (1-\delta k)\log\left(\frac{1-\delta k}{N-k}\right),$$

and the optimum is achieved for the probability distribution with

$$p(1) = \cdots = p(k) = \delta \quad \text{and} \quad p(k+1) = \cdots = p(N) = \frac{1-\delta k}{N-k}.$$

Proof. Exercises. □

Now, we can prove the main result of this lecture:

Theorem 1.2 (HSW)**.** *For any quantum channel $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$ we have*

$$C(T) = \lim_{k\to\infty} \frac{1}{k}\chi\left(T^{\otimes k}\right).$$

Proof. Using the coding schemes constructed for product codes in the previous lecture, we find that

$$\chi(T) \leqslant C(T),$$

for any quantum channel $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$. Now, we apply this equation for the quantum channel $T^{\otimes k}$ and using an exercise we obtain

$$\frac{1}{k}\chi\left(T^{\otimes k}\right) \leqslant \frac{1}{k}C(T^{\otimes k}) = C(T).$$

Finally, taking limits of subsequences shows that

$$\limsup_{k\to\infty}\frac{1}{k}\chi\left(T^{\otimes k}\right)\leqslant C(T).$$

We will now derive an upper bound on the classical capacity $C(T)$ from which the theorem will follow. For this, we assume that for every $n\in\mathbb{N}$ there exists an $m_n\in\mathbb{N}$ such that

$$R=\lim_{n\to\infty}\frac{m_n}{n},$$

and such that there is a map

$$E_n:\{0,1\}^{m_n}\to D(\mathcal{H}_A^{\otimes n}),$$

and a measurement

$$\mu_n:\{0,1\}^{m_n}\to B(\mathcal{H}_B)^{\otimes n},$$

with

$$\langle\mu(b_1,\ldots,b_{m_n}),T^{\otimes n}\circ E(b_1,\ldots,b_{m_n})\rangle_{HS}=1-\delta_n,$$

for all $b_1,\ldots,b_{m_n}\in\{0,1\}$ and such that $\lim_{n\to\infty}\delta_n=0$. To apply Holevo's theorem, we consider the uniform probability distribution $p_{A^n}\in\mathcal{P}\left(\{0,1\}^{m_n}\right)$ given by

$$p_{A^n}(b_1,\ldots,b_n)=\frac{1}{2^{m_n}},$$

and consider the ensemble

$$\{p_{A^n}(a_1,\ldots,a_{m_n}),T^{\otimes n}\circ E(a_1,\ldots,a_{m_n})\}.$$

We may now define a joined probability distribution by

$$p_{A^nB^n}(a_1,\ldots,a_{m_n},b_1,\ldots,b_{m_n})=p_{A^n}\left(a_1,\ldots,a_{m_n}\right)\langle\mu(b_1,\ldots,b_{m_n}),T^{\otimes n}\circ E(a_1,\ldots,a_{m_n})\rangle_{HS},$$

and by Holevo's theorem, we have

$$I\left(A^n:B^n\right)_{p_{A_nB_n}}\leqslant I_{\mathrm{acc}}\left(\{p_A(a_1,\ldots,a_{m_n}),T^{\otimes n}\circ E(a_1,\ldots,a_{m_n})\}\right)\leqslant\chi\left(T^{\otimes n}\right).$$

Let us estimate the mutual information on the left-hand side of the previous equation. Since p_{A^n} is uniform on $\{0,1\}^{m_n}$, we have

$$H(p_{A^n})=m_n.$$

Note that

$$\begin{aligned}p_{B^n}(b_1,\ldots,b_{m_n})&=\sum_{a_1,\ldots,a_{m_n}}\frac{1}{2^{m_n}}\langle\mu(b_1,\ldots,b_{m_n}),T^{\otimes n}\circ E(a_1,\ldots,a_{m_n})\rangle_{HS}\\&\geqslant\frac{1-\delta_n}{2^{m_n}},\end{aligned}$$

for each $(b_1,\ldots,b_{m_n})\in\{0,1\}^{m_n}$ and hence

$$p_{B^n}=(1-\delta_n)u+\delta_nq,$$

with the uniform distribution $u\in\mathcal{P}\left(\{0,1\}^{m_n}\right)$ given by $u(b_1,\ldots,b_{m_n})=1/2^{m_n}$ and some other probability distribution $q\in\mathcal{P}\left(\{0,1\}^{m_n}\right)$. By concavity of Shannon's entropy we have

$$H(p_{B^n})\geqslant(1-\delta_n)m_n.$$

Finally, note that

$$p_{A^nB^n}(b_1,\ldots,b_{m_n},b_1,\ldots,b_{m_n}) = \frac{1-\delta_n}{2^{m_n}},$$

for every $(b_1,\ldots,b_{m_n}) \in \{0,1\}^{m_n}$ and by Lemma 1.1 we conclude that

$$\begin{aligned} H(p_{A^nB^n}) &\leqslant -(1-\delta_n)\log\left(\frac{1-\delta_n}{2^{m_n}}\right) - \delta_n \log\left(\frac{\delta_n}{2^{2m_n}-2^{m_n}}\right) \\ &\leqslant (1+\delta_n)m_n + H((1-\delta_n,\delta_n)) \leqslant (1+\delta_n)m_n + 1. \end{aligned}$$

With this we compute that

$$I\left(A^n : B^n\right)_{p_{A_nB_n}} = H(p_{A^n}) + H(p_{B^n}) - H(p_{A^nB^n}) \geqslant (1-2\delta_n)m_n - 1.$$

We conclude that

$$(1-2\delta_n)\frac{m_n}{n} - \frac{1}{n} \leqslant \frac{1}{n} I\left(A^n : B^n\right)_{p_{A_nB_n}} \leqslant \frac{1}{n}\chi\left(T^{\otimes n}\right),$$

for any $n \in \mathbb{N}$. Taking limits of subsequences we conclude that

$$R \leqslant \liminf_{n\to\infty} \frac{1}{n}\chi\left(T^{\otimes n}\right).$$

By combining this with the lower bound, we find that

$$\limsup_{k\to\infty} \frac{1}{k}\chi\left(T^{\otimes k}\right) \leqslant C(T) \leqslant \liminf_{k\to\infty} \frac{1}{k}\chi\left(T^{\otimes k}\right),$$

and we conclude that

$$C(T) = \lim_{k\to\infty} \frac{1}{k}\chi\left(T^{\otimes k}\right).$$

□

2 The capacity of entanglement breaking channels

We need the following definition, which already appeared in Lecture 6:

Definition 2.1. *A linear map $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$ is called* entanglement breaking *(EB) if $C_T \in \mathrm{Sep}\left(\mathcal{H}_A, \mathcal{H}_B\right)$. We denote the set of entanglement breaking maps by $EB(\mathcal{H}_A \to \mathcal{H}_B)$.*

The following theorem contains a few equivalent characterizations of these linear maps:

Theorem 2.2 (Characterizing entanglement breaking channels)**.** *For a linear map $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$ the following are equivalent:*

1. *We have $T \in EB(\mathcal{H}_A \to \mathcal{H}_B)$.*
2. *We have*
$$\left(\mathrm{id}_R \otimes T\right)\left(X_{RA}\right) \in \mathrm{Sep}\left(\mathcal{H}_R, \mathcal{H}_B\right)$$
for every Euclidean space $\mathcal{H}_R$ and every $X_{RA} \in B(\mathcal{H}_R \otimes \mathcal{H}_A)^+$.
3. *We have the Kraus decomposition*
$$T = \sum_{n=1}^{N} \mathrm{Ad}_{K_n},$$
with $\mathrm{rk}\left(K_n\right) = 1$ for each $n \in \{1,\ldots,N\}$.

4. We have a decomposition

$$T = \sum_{l=1}^{L} \langle A_l, \cdot \rangle_{HS} B_l,$$

with $A_l \in B(\mathcal{H}_A)^+$ *and* $B_l \in B(\mathcal{H}_B)^+$.

The decomposition in the fourth point is sometimes called the Holevo representation *of the entanglement breaking map.*

Entanglement breaking quantum channels, as the name suggests, break the entanglement between any system they are applied to and any other system, and they can be thought of as a POVM measurement followed by a preparation conditioned on the measurement outcome (see point 4. in the previous theorem). The entanglement breaking channels were among the first classes of quantum channels for which the regularized Holevo information, and hence the classical capacity, could be computed. We will now discuss how this was achieved:

Theorem 2.3. *Let* $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$ *denote an entanglement breaking quantum channel and* $S : B(\mathcal{H}_C) \to B(\mathcal{H}_D)$ *any quantum channel. We have*

$$\chi(T \otimes S) = \chi(T) + \chi(S).$$

Proof. For any ensembles $\{p(x), \rho_x\}_{x \in \Sigma_1}$ with $\rho_x \in D(\mathcal{H}_A)$ and $\{q(y), \sigma_y\}_{y \in \Sigma_2}$ with $\sigma_y \in D(\mathcal{H}_C)$, we have

$$\chi(\{p(x), T(\rho_x)\}) + \chi(\{q(y), S(\sigma_y)\}) = \chi(\{p(x)q(y), T(\rho_x) \otimes S(\sigma_y)\}) \leqslant \chi(T \otimes S).$$

Hence, we have

$$\chi(T \otimes S) \geqslant \chi(T) + \chi(S).$$

For the remaining inequality, we consider an ensemble $\{p(x), \rho_x\}_{x \in \Sigma}$ with $\rho_x \in D(\mathcal{H}_A \otimes \mathcal{H}_C)$ such that

$$\chi(T \otimes S) = \chi(\{p(x), (T \otimes S)(\rho_x)\}),$$

which exists by an exercise. Since T is an entanglement breaking quantum channel, we have

$$(T \otimes \mathrm{id}_C)(\rho_x) = \sum_{y'} q'(x, y') \tau_{xy'} \otimes \eta_{xy'} = \sum_{y} q(x, y) |b_{xy}\rangle\langle b_{xy}| \otimes |c_{xy}\rangle\langle c_{xy}|,$$

for probability distributions $q'(x, \cdot)$ and $q(x, \cdot)$, quantum states $\tau_{xy'} \in D(\mathcal{H}_B)$ and $\eta_{xy'} \in D(\mathcal{H}_C)$ and some vectors $|b_{xy}\rangle \in \mathcal{H}_B$ and $|c_{xy}\rangle \in \mathcal{H}_C$ obtained by the spectral decomposition, for any $x \in \Sigma$. Next, we consider the quantum state

$$\sigma^{(x)}_{ABD} = \sum_{y} q(x, y) |y\rangle\langle y| \otimes |b_{xy}\rangle\langle b_{xy}| \otimes S(|c_{xy}\rangle\langle c_{xy}|)$$

By strong subadditivity of the von-Neumann entropy, we have

$$H(\sigma^{(x)}_{BD}) \geqslant H(\sigma^{(x)}_{ABD}) - H(\sigma^{(x)}_{AB}) + H(\sigma^{(x)}_{B}). \tag{1}$$

Let us compute the entropies appearing in the previous inequality: It is easy to see that

$$\sigma^{(x)}_{BD} = \sum_{y} q(x, y) |b_{xy}\rangle\langle b_{xy}| \otimes S(|c_{xy}\rangle\langle c_{xy}|) = (T \otimes S)(\rho_x),$$

and hence we have

$$H(\sigma^{(x)}_{BD}) = H((T \otimes S)(\rho_x)). \tag{2}$$

Next, note that

$$\begin{aligned} H(\sigma_{ABD}^{(x)}) &= H\left(\sum_y q(x,y)|y\rangle\langle y| \otimes |b_{xy}\rangle\langle b_{xy}| \otimes S\left(|c_{xy}\rangle\langle c_{xy}|\right)\right) \\ &= H(q) + \sum_y q(x,y) H\left(|b_{xy}\rangle\langle b_{xy}| \otimes S\left(|c_{xy}\rangle\langle c_{xy}|\right)\right) \\ &= H(q) + \sum_y q(x,y) H\left(S\left(|c_{xy}\rangle\langle c_{xy}|\right)\right), \end{aligned}$$

and similarly we also have

$$H\left(\sigma_{AB}^{(x)}\right) = H\left(\sigma_{A}^{(x)}\right) = H(q).$$

Combining the previous two equations shows that

$$H(\sigma_{ABD}^{(x)}) - H(\sigma_{AB}^{(x)}) = \sum_y q(x,y) H\left(S\left(|c_{xy}\rangle\langle c_{xy}|\right)\right). \tag{3}$$

Finally, we have

$$H(\sigma_{B}^{(x)}) = H\left(\sum_y q(x,y)|b_{xy}\rangle\langle b_{xy}|\right) = H\left(T\left(\mathrm{Tr}_C\left[\rho_x\right]\right)\right). \tag{4}$$

Combining (1),(2),(3) and (4) we find that

$$H\left(\left(T \otimes S\right)(\rho_x)\right) \geqslant \sum_y q(x,y) H\left(S\left(|c_{xy}\rangle\langle c_{xy}|\right)\right) + H\left(T\left(\mathrm{Tr}_C\left[\rho_x\right]\right)\right).$$

Summing over $x \in \Sigma$ shows that

$$\sum_x p(x) H\left(\left(T \otimes S\right)(\rho_x)\right) \geqslant \sum_x \sum_y p(x) q(x,y) H\left(S\left(|c_{xy}\rangle\langle c_{xy}|\right)\right) + \sum_x p(x) H\left(T\left(\mathrm{Tr}_C\left[\rho_x\right]\right)\right).$$

Now, we compute

$$\begin{aligned} \chi\left(\{p(x), (T\otimes S)(\rho_x)\}\right) &= H\left(\sum_x p(x)(T\otimes S)(\rho_x)\right) - \sum_x p(x) H\left((T\otimes S)(\rho_x)\right) \\ &\leqslant H\left(T\left(\mathrm{Tr}_C\left[\sum_x p(x)(\rho_x)\right]\right)\right) + H\left(S\left(\mathrm{Tr}_A\left[\sum_x p(x)(\rho_x)\right]\right)\right) \\ &\quad - \sum_x \sum_y p(x) q(x,y) H\left(S\left(|c_{xy}\rangle\langle c_{xy}|\right)\right) - \sum_x p(x) H\left(T\left(\mathrm{Tr}_C\left[\rho_x\right]\right)\right) \\ &= \chi\left(\{p(x), T\left(\mathrm{Tr}_C\left[\rho_x\right]\right)\}\right) + \chi\left(\{p(x)q(x,y), S\left(|c_{xy}\rangle\langle c_{xy}|\right)\}\right), \end{aligned}$$

since

$$\sum_x \sum_y p(x) q(x,y) |c_{xy}\rangle\langle c_{xy}| = \mathrm{Tr}_A\left[\sum_x p(x)\rho_x\right].$$

We conclude that

$$\chi\left(T \otimes S\right) = \chi\left(\{p(x), (T\otimes S)(\rho_x)\}\right) \leqslant \chi(T) + \chi(S).$$

This finishes the proof.

□

As a corollary of the previous theorem and the HSW theorem we have:

Corollary 2.4. *For any entanglement breaking channel $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$ we have*

$$C(T) = \chi(T).$$

3 Other examples where the Holevo information is additive

Without proof, we mention the following classes of quantum channels $T : B(\mathcal{H}) \to B(\mathcal{H})$ for which we have

$$\chi\left(T \otimes S\right) = \chi\left(T\right) + \chi\left(S\right),$$

for any quantum channel $S : B(\mathcal{H}') \to B(\mathcal{H}'')$. For each of these classes we we have $C(T) = \chi(T)$.

Theorem 3.1 (Additivity examples)**.** *The Holevo quantity is additive if T is*

- *a unital qubit channel.*
- *a depolarizing channel, i.e., such that $T : B(\mathcal{H}) \to B(\mathcal{H})$ is given by*

  $$T(X) = (1-p)\operatorname{Tr}\left[X\right]\frac{\mathbb{1}_{\mathcal{H}}}{\dim(\mathcal{H})} + pX,$$

  for any $X \in B(\mathcal{H})$.
- *a Schur multiplier channel, i.e., such that $T : B(\mathcal{H}) \to B(\mathcal{H})$ is given by*

  $$T(X) = A \odot X,$$

  for any $X \in B(\mathcal{H})$, where $A \in B(\mathcal{H})^{+}$ and where $\odot$ denotes the entrywise product in the computational basis.