VU Resource Theories and Thermodynamics
Exercise sheet 3

Lorenz Hummer

April 2, 2020

3 Exercises - Week 3

3.1 Quantum operations

(i). The fact that unitary maps are CPTP follows from the fact that unitary conjugation leaves the eigenvalues of an operator invariant. This can be shown by means of the characteristic polynomial:

$$\begin{aligned}\det\Big[U\rho U^\dagger - \lambda\mathbb{I}\Big] &= \det\Big[U\rho U^\dagger - UU^\dagger\lambda\mathbb{I}\Big] \\ &= \det\Big[U(\rho-\lambda\mathbb{I})U^\dagger\Big] \\ &= \det\Big[UU^\dagger\Big]\det[\rho-\lambda\mathbb{I}] \\ &= \det[\rho-\lambda\mathbb{I}]\end{aligned}$$

where in the third equality I have used that $\det(AB) = \det(A)\det(B)$.

Since the trace is just the sum of the eigenvalues (with multiplicities), this means that the trace is preserved. This can also be shown by the cyclic property of the trace:

$$\mathrm{tr}\Big[U\rho U^\dagger\Big] = \Bigg[\underbrace{U^\dagger U}_{\mathbb{I}}\rho\Bigg] = \mathrm{tr}[\rho] = 1$$

Complete positivity means that if $\rho \geq 0$, a map $\Lambda : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ satisfies

$$(\Lambda \otimes \mathbb{I}_k)[\rho] \geq 0$$

for arbitrary integer k. But $U \otimes \mathbb{I}$ is also unitary, since

$$(U\otimes\mathbb{I})(U\otimes\mathbb{I})^\dagger = (U\otimes\mathbb{I})(U^\dagger\otimes\mathbb{I}) = UU^\dagger\otimes\mathbb{I} = \mathbb{I}$$

Hence, the eigenvalues are again preserved. Since a symmetric operator is positive semidefinite if and only if all eigenvalues are nonnegative, it follows that unitaries are completely positive.

(ii). The transpose map $\mathcal{T} : \rho \to \rho^T$ is an example for a linear map that is positive but not completely positive. Since the trace of a matrix is just the sum of its diagonal elements, transposition obviously leaves the trace invariant. Transposition also preserves the spectrum:

$$\det\left[\rho^T - \lambda\mathbb{I}\right] = \det\left[(\rho - \lambda\mathbb{I})^T\right] = \det[\rho - \lambda\mathbb{I}]$$

because $\det[A] = \det[A^T]$.

Hence it maps every positive operator onto a positive operator. To show that the transposition is not completely positive, it suffices to consider the two-qubit state

$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

The corresponding density matrix is

$$|\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|)$$

Applying the partial transpose, i.e. the transpose on only one of the two qubits (say qubit A), gives:

$$\rho' := (\mathcal{T} \otimes \mathbb{I})[\rho] = \frac{1}{2}(|00\rangle\langle 00| + |10\rangle\langle 01| + |01\rangle\langle 10| + |11\rangle\langle 11|)$$

However

$$|\Psi^-\rangle := \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

is an eigenvector of ρ' with negative eigenvalue:

$$\rho'|\Psi^-\rangle = \frac{1}{2\sqrt{2}}(|10\rangle - |01\rangle) = -\frac{1}{2}|\Psi^-\rangle$$

Hence, $\mathcal{T}$ is not completely positive.

(iii). Given an orthonormal basis $\{|i\rangle\}_{i=0}^d$ in which the density matrix of the initial state ρ is diagonal, a possible choice of Kraus operators for the preparation of a pure state $|\psi\rangle$ is given by $\{A_i\}_{i=1}^d$, where

$$A_i := |\psi\rangle\langle i|$$

and d is the dimension of the corresponding Hilbert space. Then we have that

$$\sum_i A_i \rho A_i^\dagger = \sum_i |\psi\rangle\langle i| \left(\sum_{j=1}^{d} \lambda_j |j\rangle\langle j| \right) |i\rangle\langle\psi| = |\psi\rangle\langle\psi|$$

$$\sum_i A_i^\dagger A_i = \sum_i |i\rangle \langle\psi|\psi\rangle \langle i| = \sum_i |i\rangle\langle i| = \mathbb{I}$$

The environment can always be chosen to be in a pure state, say $|0\rangle\langle 0|_{env}$. In that case, the unitary U that implements the quantum operation on the joint state of the input and the ancilla can be constructed by the rule that[1] (cf. [1], p. 268)

$$A_i = \langle i|U|0\rangle$$

That is, the Kraus operators specify the first column block of the matrix. The other elements have to be chosen in such a way that it becomes a unitary. In the present example, this means that U takes the following form for a 2-dimensional system:

$$U = \begin{pmatrix} |\psi\rangle\langle 0| & |\psi^\perp\rangle\langle 0| \\ |\psi\rangle\langle 1| & -|\psi^\perp\rangle\langle 1| \end{pmatrix}$$

where $\langle\psi|\psi^\perp\rangle = 0$. This matrix has the Kraus operators in its first "column" and is unitary because

$$\begin{aligned} UU^\dagger &= \begin{pmatrix} |\psi\rangle\langle 0| & |\psi^\perp\rangle\langle 0| \\ |\psi\rangle\langle 1| & -|\psi^\perp\rangle\langle 1| \end{pmatrix} \cdot \begin{pmatrix} |0\rangle\langle\psi| & |1\rangle\langle\psi| \\ |0\rangle\langle\psi^\perp| & -|1\rangle|\psi^\perp\rangle \end{pmatrix} \\ &= \begin{pmatrix} |\psi\rangle\langle\psi| + |\psi^\perp\rangle\langle\psi^\perp| & 0 \\ 0 & |\psi\rangle\langle\psi| + |\psi^\perp\rangle\langle\psi^\perp| \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix} \end{aligned}$$

It implements the quantum channel via

$$|\psi\rangle\langle\psi| = \mathrm{tr}_{env}\Big[U(\rho \otimes |0\rangle\langle 0|_{env})U^\dagger\Big] = \sum_{i=1}^{2} \langle i|U|0\rangle\, \rho\, \langle 0|U^\dagger|i\rangle$$

[1] $\langle i|U|0\rangle$ is just shorthand for $\left[\mathbb{I} \otimes \langle i|_{env}\right] U \left[\mathbb{I} \otimes |0\rangle_{env}\right]$

As for the decoherence, one can choose the Kraus operators

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = B_1^\dagger$$

$$B_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = B_2^\dagger$$

Then we have that

$$\begin{aligned} \sum_i B_i \rho B_i^\dagger &= \left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ b^* & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ b^* & c \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right] \\ &= \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & c \end{pmatrix} \\ &= \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix} \end{aligned}$$

and that

$$B_1^\dagger B_1 + B_2^\dagger B_2 = \mathbb{I}$$

A unitary U that implements the quantum operation when an ancilla in the state $|0\rangle\langle 0|_{env}$ is added is given by

$$U = \begin{pmatrix} |0\rangle\langle 0| & |1\rangle\langle 1| \\ |1\rangle\langle 1| & |0\rangle\langle 0| \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

For a random replacement with a diagonal state $\mathrm{diag}(\gamma_1, \ldots, \gamma_n)$, which occurs with probaility p, one can choose (in the basis $\{|i\rangle\}_{i=0}^{n-1}$, in which the input state is diagonal) the following set of Kraus operators

$$\{C_{ij}\}_{i,j=0}^{n-1} \cup \sqrt{1-p}\,\mathbb{I}$$

where

$$C_{ij} := \sqrt{p\,\gamma_{i+1}}|i\rangle\langle j|$$

Specifically, for $n = 2$:

$$C_{00} = \sqrt{p\,\gamma_1}|0\rangle\langle 0|$$
$$C_{01} = \sqrt{p\,\gamma_1}|0\rangle\langle 1|$$
$$C_{10} = \sqrt{p\,\gamma_2}|1\rangle\langle 0|$$
$$C_{11} = \sqrt{p\,\gamma_2}|1\rangle\langle 1|$$

Then,

$$\begin{aligned}\sum_{ij} C_{ij}\rho C_{ij}^\dagger &= \sum_{ij} C_{ij}(\lambda|0\rangle\langle 0| + (1-\lambda)|1\rangle\langle 1|)C_{ij}^\dagger \\ &= p\gamma_1(\lambda|0\rangle\langle 0| + (1-\lambda)|0\rangle\langle 0|) + p\gamma_2(\lambda|1\rangle\langle 1| + (1-\lambda)|1\rangle\langle 1|) \\ &= p\,\mathrm{diag}(\gamma_1,\gamma_2)\end{aligned}$$

and

$$\sum_{ij} C_{ij}^\dagger C_{ij} = p(\gamma_1+\gamma_2)(|0\rangle\langle 0| + |1\rangle\langle 1|) = p\mathbb{I}$$

provided that $\gamma_1 + \gamma_2 = 1$. Adding the Kraus operator $\sqrt{1-p}\,\mathbb{I}$ then gives the new state

$$\rho' = p\,\mathrm{diag}(\gamma_1,\gamma_2) + (1-p)\rho$$

3.2 Unital maps and noisy operations

(i). Quantum noisy operations are unital because for $\rho = \mathbb{I}_A/d_A$,

$$\mathcal{E} : \rho \mapsto \sigma = \mathrm{tr}_B\left[U\left(\frac{\mathbb{I}_A}{d_A}\otimes\frac{\mathbb{I}_B}{d_B}\right)U^\dagger\right] = \frac{1}{d_A d_B}\mathrm{tr}_B[\underbrace{UU^\dagger}_{\mathbb{I}_{AB}}] = \frac{1}{d_A}\mathbb{I}_A$$

Intuitively, you can't create a less mixed state by a noisy operation (the entropy has to be non-decreasing).

(ii). In the classical case, a noisy map has the form

$$\mathcal{D} : \vec{r} \mapsto \vec{s} := [\pi(\vec{r}\otimes\vec{\gamma}_{d_1})]_A$$

where $\vec{\gamma}_d := (1/d,\ldots,1/d)^T \in \mathbb{R}^d$ and $[...]_A$ is the marginal distribution with dimension d.

We first have to show that classical noisy operations must be uniform-preserving stochastic maps. As for the former property, we have that for $\vec{r} := \vec{\gamma}_d$:

$$\mathcal{D}(\vec{r}) = [\pi(\vec{r}\otimes\vec{\gamma}_{d_B})]_A = \left[\pi\left(\frac{1}{dd_B},\ldots,\frac{1}{dd_B}\right)^T\right]_A = \left[\left(\frac{1}{dd_B},\ldots,\frac{1}{dd_B}\right)^T\right]_A$$

Computing the marginal means summing over the d_B-blocks, i.e.

$$[\mathcal{D}(\vec{r})]_i = \sum_{j=1+(i-1)d_B}^{id_B}\left[\left(\frac{1}{dd_B},\ldots,\frac{1}{dd_B}\right)^T\right]_j = d_B\frac{1}{d_B}(\vec{\gamma}_d)_i = (\vec{\gamma}_d)_i$$

It is also clear that they are stochastic maps, since they are exactly those that map probability vectors into probability vectors as was shown in the previous exercises. If $\vec{r}$ is a probability vector, then so is $\mathcal{D}(\vec{r})$, because

$$[\mathcal{D}(\vec{r})]_i = \sum_{j=1+(i-1)d_B}^{id_B}\left[\pi\left(\underbrace{\frac{r_1}{d_B}}_{\geq 0},\ldots,\underbrace{\frac{r_1}{d_B}}_{\geq 0},\underbrace{\frac{r_2}{d_B}}_{\geq 0},\ldots,\underbrace{\frac{r_2}{d_B}}_{\geq 0},\ldots,\underbrace{\frac{r_d}{d_B}}_{\geq 0},\ldots,\underbrace{\frac{r_d}{d_B}}_{\geq 0}\right)^T\right]_j \geq 0$$

$$\sum_{i=1}^{d}[\mathcal{D}(\vec{r})]_i = d_B\sum_{i=1}^{d}\frac{r_i}{d_B} = \sum_{i=1}^{d} r_i = 1$$

(iii). (cf. [2]). It remains to be shown that every uniform-preserving stochastic map D can be implemented to arbitrary accuracy as a noisy operation. For that, we use the results of last week's exercises to argue that D is bistochastic if the input and output dimensions are equal (because they are the subset of stochastic matrices that preserve the uniform distribution) and that, due to Birkhoff's theorem, D is a mixture of permutations $D = \sum_i p_i P_i$. It remains to be shown that any such mixture can be cast as a classical noisy operation.

To that end, prepare a sufficiently large ancilla of size N and split it into subsets of size N_i such that their relative sizes correspond to the probabilities with which we apply the respective permutations, i.e. $p_i = N_i/N$. Assume that D is the sum of l permutations. Then we have that

$$(\vec{r}\otimes\vec{\gamma}_N) = \left(\underbrace{\frac{r_1}{N}}_{N_1\text{ times}},\underbrace{\frac{r_1}{N}}_{N_2\text{ times}},\ldots,\underbrace{\frac{r_1}{N}}_{N_l\text{ times}},\underbrace{\frac{r_2}{N}}_{N_1\text{ times}},\ldots,\underbrace{\frac{r_2}{N}}_{N_l\text{ times}},\ldots,\underbrace{\frac{r_d}{N}}_{N_1\text{ times}},\ldots,\underbrace{\frac{r_d}{N}}_{N_l\text{ times}}\right)^T$$

If we now permute the elements of each individual ancilla subset among each other, i.e. apply permutation P_1 to the elements

$$\Bigg(\underbrace{r_1}_{N_1 \text{ times}}, \ldots, \underbrace{r_d}_{N_1 \text{ times}}\Bigg)$$

in ancilla subset 1, P_2 to the elements

$$\Bigg(\underbrace{r_1}_{N_2 \text{ times}}, \ldots, \underbrace{r_d}_{N_2 \text{ times}}\Bigg)$$

in ancilla subset 2 etc. up to P_l, this can also be regarded as a particular *global* permutation on the joint system, i.e.

$$\pi(\vec{r}\otimes\vec{\gamma}_N) = \frac{1}{N}\Bigg(\underbrace{r_{P_1(1)}}_{N_1 \text{ times}}, \underbrace{r_{P_2(1)}}_{N_2 \text{ times}}, \ldots, \underbrace{r_{P_l(1)}}_{N_l \text{ times}}, \underbrace{r_{P_1(2)}}_{N_1 \text{ times}}, \ldots, \underbrace{r_{P_l(2)}}_{N_l \text{ times}}, \ldots, \underbrace{r_{P_1(d)}}_{N_1 \text{ times}}, \ldots, \underbrace{r_{P_l(d)}}_{N_l \text{ times}}\Bigg)^T$$

Now marginalizing, i.e. removing the ancilla, yields:

$$\begin{aligned}
[\pi(\vec{r}\otimes\vec{\gamma}_N)]_A &= \frac{1}{N}\left(\sum_{i_1=1}^{l} N_{i_1} r_{P_{i_1}(1)}, \sum_{i_2=1}^{l} N_{i_2} r_{P_{i_2}(2)}, \ldots, \sum_{i_d=1}^{l} N_{i_d} r_{P_{i_d}(d)}\right)^T \\
&= \frac{N_1}{N} P_1 \vec{r} + \cdots + \frac{N_l}{N} P_l \vec{r} \xrightarrow[N\to\infty]{} \sum_{i=1}^{l} p_i P_i \vec{r}
\end{aligned}$$

We have thus constructed an arbitrary mixture of permutations via a noisy classical operation. As mentioned above, because any uniform-preserving stochastic matrix has this form, any such matrix can be implemented as noisy classical operation (to arbitrary accuracy).

(iv). (cf. [2]). When input and output dimensions d_{in} and d_{out} are not equal, a classical noisy operation may be defined as a map $\mathcal{D} : V_{in} \to V_{out}$ that allows a decomposition of the form

$$\mathcal{D}[\vec{p}] = [\pi(\vec{p}\otimes\vec{\gamma}_{d_A})]_{A'}$$

This is very similar to the definition with $d_{in} = d_{out}$ with the exception that the marginalization is not performed over the ancilla space of the input state but a state space A' of dimension $d_{A'}$ that is complementary to the output state space in the sense that $d_{out}d_{A'} = d_{in}d_A$. As for the uniform-preservation property, we require that the uniform state of dimension d_{in} be mapped onto the uniform state of dimension d_{out}, $\mathcal{D}[\vec{\gamma}_{d_{in}}] = \vec{\gamma}_{d_{out}}$

If the dimensions d_{in} and d_{out} are unequal, one can add ancillary systems to the input and the output such that the new composite systems are of equal dimension: $d_{out}d_2 = d_{in}d_1$. Then, one can define a $d_2 \times d_1$-matrix D_0 with $D_{0,ij} = 1/d_2 \,\forall i, j$, which maps any d_1-dimensional probability distribution onto the uniform distribution of dimension d_2, since

$$\sum_j D_{ij} p_j = \frac{1}{d_2} \sum_j p_j = \frac{1}{d_2} \,\forall i$$

In particular, it is uniform-preserving and stochastic (because $D_{0,ij} \geq 0$ and $\sum_i D_{0,ij} = d_2(1/d_2) = 1$). If two matrices A and B are stochastic, then so is their tensor product $A \otimes B$. This is seen as follows:

$$A \otimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1m}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nm}B \end{pmatrix}$$

Hence, $(A \otimes B)_{ij} \geq 0 \,\forall i, j$ and

$$\sum_i (A \otimes B)_{ij} = \sum_{ik} A_{ij} B_{kj} = \underbrace{\left(\sum_i A_{ij}\right)}_{1} \underbrace{\left(\sum_k B_{kj}\right)}_{1} = 1$$

$D \otimes D_0$ is hence stochastic. It is also uniform-preserving, because

$$\sum_j (D \otimes D_0)_{ij} \frac{1}{d_{in}d_1} = \frac{1}{d_{in}d_1} d_1 \frac{1}{d_2} \underbrace{\sum_j D_{ij}}_{d_{in}/d_{out}} = \frac{1}{d_2 d_{out}}$$

Therefore, $D \otimes D_0$ is doubly stochastic and can hence be written as a mixture of permutations. Since adjoining a uniform state of state of dimension d_1, implementing a mixture of permutations and marginalizing over the ancilla system of dimension d_2 are all noisy operations, the total one is as well.

3.3 Classical and quantum noisy operations

(i). Given the eigendecomposition of two density matrices ρ and σ

$$\rho = \sum_{i=1}^{d_{in}} \lambda_i(\rho) |\phi_i\rangle\langle\phi_i|$$

$$\sigma = \sum_{j=1}^{d_{out}} \lambda_j(\sigma) |\psi_j\rangle\langle\psi_j|$$

we first need to show that the existence of a noisy qunatum operation $\mathcal{E} : \rho \mapsto \sigma$ implies that there exists a noisy classical operation $\lambda(\rho) \mapsto \lambda(\sigma)$. Expanding $\sigma = \mathcal{E}(\rho)$, we obtain:

$$\sum_{j=1}^{d_{out}} \lambda_j(\sigma)|\psi_j\rangle\langle\psi_j| = \mathrm{tr}_B \left[U\rho \otimes \gamma_B U^\dagger \right] = \mathrm{tr}_B \left[U \left(\sum_{i=1}^{d_{in}} \lambda_i(\rho)|\psi_j\rangle\langle\psi_j| \right) \otimes \gamma_B U^\dagger \right]$$

where $\gamma_B \coloneqq \mathbb{I}/d$ is the maximally mixed state. Now, to find $\lambda_k(\sigma)$, compute:

$$\begin{aligned} \lambda_k(\sigma) &= \langle\psi_k|\sigma|\psi_k\rangle \\ &= \langle\psi_k|\mathrm{tr}_B \left[U \left(\sum_{i=1}^{d_{in}} \lambda_i(\rho)|\phi_i\rangle\langle\phi_i| \right) \otimes \gamma_B U^\dagger \right] |\psi_k\rangle \\ &= \sum_{i=1}^{d_{in}} \lambda_i(\rho) \, \langle\psi_k| \underbrace{\mathrm{tr}_B \left[U|\phi_i\rangle\langle\phi_i| \otimes \gamma_B U^\dagger \right]}_{\mathcal{E}(|\phi_i\rangle\langle\phi_i|)} |\psi_k\rangle \end{aligned}$$

Hence,

$$\lambda_j(\sigma) = \sum_i D_{ji}\lambda_i(\rho)$$

where

$$D_{ji} = \langle\psi_j|\mathcal{E}(|\phi_i\rangle\langle\phi_i|)|\psi_j\rangle$$

It can be implemented to arbitrary accuracy as a noisy operation because it is a uniform-preserving stochastic map. To see this, suppose $\rho = \mathbb{I}/d_{in}$. Then,

$$\sum_i D_{ji}\lambda_i(\rho) = \frac{1}{d_{in}} \langle\psi_j| \frac{d_{in}}{d_{out}} \mathbb{I} |\psi_j\rangle = \frac{1}{d_{out}}$$

where the first equality comes from the fact that noisy quantum operations are unital. The above equation implies that $\sigma = \mathbb{I}_{out}/d_{out}$, hence D is uniform preserving.

We also have that

$$\sum_j D_{ji} = \sum_j \langle\psi_j|\mathcal{E}(|\phi_i\rangle\langle\phi_i|)|\psi_j\rangle = \mathrm{tr}[\mathcal{E}(|\phi_i\rangle\langle\phi_i|)] = \mathrm{tr}[|\phi_i\rangle\langle\phi_i|] = \langle\phi_i|\phi_i\rangle = 1$$

where the third equality is justified by the fact that noisy-quantum operations are trace-preserving. Together with the fact that $D_{ij} \geq 0\,\forall i,j$, the last equation demonstrates that D is a stochastic matrix.

(ii). To show the converse, assume that D is a classical noisy operation such that $\lambda(\sigma) = D\lambda(\rho)$. Define

$$\mathcal{E}(\rho) \coloneqq \sum_{k,j} D_{kj}|\psi_k\rangle \langle\phi_j|\rho|\phi_j\rangle \langle\psi_k|$$

First, we can convince ourselves that $\mathcal{E}(\rho) = \sigma$:

$$\mathcal{E}(\rho) = \sum_k \left(\sum_j D_{kj} \underbrace{\langle\phi_j|\rho|\phi_j\rangle}_{\lambda_j(\rho)} \right) |\psi_k\rangle\langle\psi_k| = \sum_k \lambda_k(\sigma)|\psi_k\rangle\langle\psi_k| = \sigma$$

Using the results from the previous exercise, we can express D in terms of a permutation R such that

$$D_{kj} = \frac{1}{d_1}\sum_{li} R_{kl,ji}$$

Now, let $\{|u_i\rangle\}$ be an orthonormal system for the ancilla system B and $\{|v_i\rangle\}$ one for the system B' that we are tracing over. When we add those systems, we obtain the expression

$$\begin{aligned}
\mathcal{E}(\rho) &= \mathrm{tr}_{B'}\left[\sum_{kj} D_{kj}|\psi_k\rangle|v_l\rangle \langle\phi_j|\langle u_i|\rho\otimes\mathbb{I}|u_i\rangle|\phi_j\rangle \langle v_l|\langle\psi_k|\right] \\
&= \mathrm{tr}_{B'}\left[\sum_{ijkl} R_{kl,ji}|\psi_k\rangle|v_l\rangle \langle\phi_j|\langle u_i|\rho\otimes\frac{\mathbb{I}}{d_1}|u_i\rangle|\phi_j\rangle \langle v_l|\langle\psi_k|\right] \\
&= \mathrm{tr}_{B'}\left[U\left(\rho\otimes\frac{\mathbb{I}}{d_1}\right)U^\dagger\right]
\end{aligned}$$

where $\frac{\mathbb{I}}{d_1} = \gamma$ and, using the fact that $\rho\otimes\mathbb{I}$ is diagonal in the basis $\{|\phi_j\rangle|u_i\rangle\}$, $U = \sum_{ijkl} R_{kl,ji}|\psi_k\rangle|v_l\rangle\langle\phi_j|\langle u_i|$.

3.4 Schur-convex functions and Rényi entropies

(i). In order to show that all functions of the type

$$F(\vec{p}) = \sum_{i=1}^{n} f(p_i) \qquad\qquad f : \mathbb{R} \to \mathbb{R} \text{ convex}$$

are Schur convex, we need to demonstrate that

$$\vec{p} \succ \vec{q} \implies \sum_{i=1}^{n} f(p_i) \geq \sum_{i=1}^{n} f(q_i)$$

From last week's exercises, we know that

$$\vec{p} \succ \vec{q} \iff \exists B \text{ bistochastic s.t. } B\vec{p} = \vec{q}$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^{n} f(q_i) &= \sum_{i=1}^{n} f\left(\sum_{j=1}^{n} B_{ij} p_j\right) \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij} f(\vec{p}_j) \\
&= \sum_{j=1}^{n} \underbrace{\left(\sum_{i=1}^{n} B_{ij}\right)}_{1} f(p_j) \\
&= \sum_{j=1}^{n} f(p_j)
\end{aligned}$$

The inequality above follows from the fact that f is convex. In fact, the definition that $F : \mathbb{R} \to \mathbb{R}$ is convex if $F(ta+(1-t)b) \leq tF(a)+(1-t)F(b)\ \forall a, b \in \mathbb{R}, t \in [0, 1]$ immediately implies that more generally,

$$F\left(\sum_j p_j a_j\right) \leq \sum_j p_j f(a_j) \quad \forall a_j \in \mathbb{R}, p_j \geq 0, \sum_j p_j = 1$$

This is proven by induction. For $n = 1$, this relation is trivial, for $n = 2$ it is just the definition of convexity. As for the induction step, we can write

$$
\begin{aligned}
F\left(\sum_{j=1}^{n+1} p_j a_j\right) &= F\left(\sum_{j=1}^{n} p_j a_j + p_{n+1}a_{n+1}\right) \\
&= F\left((1-p_{n+1})\sum_{j=1}^{n} \frac{p_j}{1-p_{n+1}} a_j + p_{n+1}a_{n+1}\right) \\
&\leq (1-p_{n+1})F\left(\sum_{j=1}^{n} \frac{p_j}{1-p_{n+1}} a_j\right) + p_{n+1}F(a_{n+1})
\end{aligned}
$$

(ii). The Rényi entropies are defined as

$$
H_\alpha(\vec{p}) \coloneqq -\frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_i^\alpha\right)
$$

In the limits, we obtain the following expressions:

For $\alpha \to 0$:

$$
\begin{aligned}
\lim_{\alpha\to 0^+} -\frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_i^\alpha\right) &= \log\left(\sum_{i=1}^{n} \lim_{\alpha\to 0^+} p_i^\alpha\right) \\
&= \log\left(\sum_{i=1}^{k} 1\right) \\
&= \log \operatorname{rank} p
\end{aligned}
$$

where k is the number of non-zero entries of $\vec{p}$ and I have used the fact that $\lim_{\alpha\to 0^+} 0^\alpha = 0$.

For $\alpha \to 1$:

$$
\begin{aligned}
\lim_{\alpha\to 1} -\frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_i^\alpha\right) &= -\lim_{\alpha\to 1} \frac{\frac{\mathrm{d}}{\mathrm{d}\alpha}\log(\sum_{i=1}^{n} p_i^\alpha)}{\frac{\mathrm{d}}{\mathrm{d}\alpha}(\alpha-1)} \\
&= -\lim_{\alpha\to 1} \underbrace{\frac{1}{\sum_{i=1}^{n} p_i^\alpha}}_{\to 1} \sum_{i=1}^{n} p_i^\alpha \log p_i \\
&= -\sum_{i=1}^{n} p_i \log p_i
\end{aligned}
$$

where the first equality is justified by L'Hospital's rule.

For the limit $\alpha \to \infty$, we can use the fact that

$$p_{max}^{\alpha} \leq \sum_{i=1}^{n} p_i^{\alpha} \leq n p_{max}^{\alpha}$$

Then,

$$\frac{1}{1-\alpha} \log(n p_{max}^{\alpha}) \leq H_{\alpha}(\vec{p}) \leq \frac{1}{1-\alpha} \log(p_{max}^{\alpha})$$

and hence,

$$\frac{1}{1-\alpha}(\log n + \alpha \log p_{max}) \leq H_{\alpha}(\vec{p}) \leq \frac{1}{1-\alpha} \alpha \log p_{max}$$

Since

$$\lim_{\alpha \to \infty} \frac{\log n}{1-\alpha} = 0$$
$$\lim_{\alpha \to \infty} \frac{\alpha}{1-\alpha} = -1$$

this implies that

$$\lim_{\alpha \to \infty} H_{\alpha}(\vec{p}) = -\log p_{max}$$

(iii). In order to prove that the Rényi entropies are Schur-concave, we can use the previously proven statement that all functions of the type

$$F(\vec{p}) = \sum_{i=1}^{n} f(p_i) \quad f : \mathbb{R} \to \mathbb{R} \text{ convex}$$

are Schur-convex.

First, consider the case $0 \leq \alpha < 1$: Since p_i^{α} is concave for $0 \leq \alpha < 1$ and $p_i > 0$ and a sum of concave functions is still concave, letting $f(p_i) := p_i^{\alpha}$ implies that $F(\vec{p}) = \sum_{i=1}^{n} f(p_i)$ is Schur-concave. Now, $1/(1-\alpha) \log x$ is monotonically increasing for $0 \leq \alpha < 1$. But a monotonically increasing function of a Schur-concave function is Schur-concave, as is easily demonstrated by the fact that if $F(\vec{p}) \leq F(\vec{q})$ for $\vec{p} \succ \vec{q}$, then $\Phi[F(\vec{p})] \leq \Phi[F(\vec{q})]$ for $\vec{p} \succ \vec{q}$ and $\Phi(x) \geq \Phi(y)$ for all $x > y$. Hence, H_{α} is Schur-concave for $0 \leq \alpha < 1$.

In case $\alpha > 1$, the argument is essentially the same. Now, p_i^{α} is convex, which implies that $F(\vec{p})$ is Schur-convex. But $1/(1-\alpha) \log x$ is monotonically decreasing for $\alpha > 1$ and a monotonically decreasing function of a Schur-convex function is Schur-concave.

For $\alpha = 1$, the Rényi entropy becomes the Shannon entropy, as shown in the previous exercise. Since $x\log(x)$ is convex, $-F(\vec{p}) = -\sum_{i=1}^{n} p_i \log p_i$ is Schur-concave, hence $H_1(\vec{p})$ is Schur-concave.

(iv). Since all Rényi entropies are Schur-concave, we know that $\vec{p} \succ \vec{q}$ implies that $H_\alpha(\vec{q}) \geq H_\alpha(\vec{p})\ \forall\alpha$. Because $\vec{p}_{pure} := (1, 0, \ldots, 0)^T \succ \vec{p} \succ (1/n, \ldots, 1/n)^T =: \vec{p}_{mix}\ \forall\vec{p}$, as shown in last week's exercises, we have[2]

$$H_\alpha(\vec{p}_{mix}) \geq H_\alpha(\vec{p}) \geq H_\alpha(\vec{p}_{pure})\ \forall\alpha$$

Now, we just have to calculate $H_\alpha(\vec{p}_{mix})$ and $H_\alpha(\vec{p}_{pure})$:

$$\begin{aligned} H_\alpha(\vec{p}_{mix}) &= -\frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_{mix,i}^{\alpha}\right) \\ &= -\frac{1}{\alpha-1}\log\left(\frac{1}{n^\alpha}\sum_{i=1}^{n} 1\right) \\ &= -\frac{1}{\alpha-1}(1-\alpha)\log n = \log n \end{aligned}$$

$$\begin{aligned} H_\alpha(\vec{p}_{pure}) &= -\frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_{pure,i}\right) \\ &= -\frac{1}{\alpha-1}\log(1) \\ &= 0 \end{aligned}$$

Both results are independent of α, so by the argument above, $0 \leq H_\alpha(\vec{p}) \leq \log n$ for any α, $\vec{p}$

[2] We can obviously take $\vec{p}_{pure} := (1, 0, \ldots, 0)^T$ w.l.o.g, because the expression for the entropies contains only the sum of the probability vector components

3.5 Data-processing inequality

(i). To demonstrate that if $X \to Y \to Z$ is a Markov chain, then so is $Z \to Y \to X$, expand $P(X = x, Y = y, Z = z)$:

$$\begin{aligned}
P(X = x, Y = y, Z = z) &= P(Z = z|X = x, Y = y)P(X = x, Y = y) \\
&= P(Z = z|Y = y)P(Y = y|X = x)P(X = x) \\
&= P(Y = y|Z = z)\frac{P(Z = z)}{P(Y = y)}P(X = x|Y = y)\frac{P(Y = y)}{P(X = x)}P(X = x) \\
&= P(X = x|Y = y)P(Y = y|Z = z)P(Z = z)
\end{aligned}$$

where in the second equality I have used that $X \to Y \to Z$ is a Markov chain and the third equality comes from Bayes' theorem:

$$P(A|B)P(B) = P(B|A)P(A)$$

(ii). Given a Markov chain $X \to Y \to Z$, one can prove $I(X;Y) \geq I(X;Z)$, where

$$I(A;B) := H(A) - H(A|B)$$

and

$$H(A|B) = H(AB) - H(B)$$

as follows:

$$\begin{aligned}
I(X;Y) - I(X;Z) &= H(X) - H(X|Y) - [H(X) - H(X|Z)] \\
&= H(X|Z) - H(X|Y) \\
&= H(X|Z) - H(X|YZ) \\
&= H(XZ) - H(Z) - [H(XYZ) - H(YZ)] \\
&= H(XZ) + H(YZ) - H(XYZ) - H(Z) \\
&\geq 0
\end{aligned}$$

where the third equality arises because $X \to Y \to Z$ (and hence also $Z \to Y \to X$) is a Markov chain and the inequality is just the strong subadditivity of the Shannon entropy.

References

[1] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement.* Cambridge: Cambridge University Press, 2006.

[2] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, "The resource theory of informational nonequilibrium in thermodynamics," *Phys. Rep.*, vol. 583, pp. 1–58, 2015.