QIT Revision Class

University of Cambridge

Part III, Lent 2018

Problem 1.

1. Define the trace distance $D(\rho, \sigma)$ between two states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ and prove that it can be expressed in the form:
$$D(\rho,\sigma) = \frac{1}{2}(\operatorname{tr} Q + \operatorname{tr} R)$$
where Q and R are suitably defined positive semi-definite operators in $\mathcal{B}(\mathcal{H})$.
2. Using the above identity, prove that
$$D(\rho,\sigma) = \max_P \operatorname{tr}(P(\rho - \sigma))$$
where the maximisation is over all projection operators $P \in \mathcal{B}(\mathcal{H})$.
3. Further, prove that
$$D(\rho,\sigma) = \max_T \operatorname{tr}(T(\rho - \sigma))$$
where the maximisation is over all positive semi-definite operators $T \in \mathcal{B}(\mathcal{H})$ with eigenvalues less than or equal to unity.
4. Let ρ be a quantum state and Λ be a linear completely positive trace-preserving map. Prove that
$$F_e(\rho, \Lambda) \le (F(\rho, \Lambda(\rho)))^2$$
where $F_e(\rho, \Lambda)$ denotes the entanglement fidelity, and $F(\rho, \Lambda(\rho))$ denotes the fidelity of the states ρ and $\Lambda(\rho)$.

Solution.

1. The trace distance is defined as
$$D(\rho,\sigma) = \frac{1}{2}\|\rho - \sigma\|_1$$
where the 1-norm of a matrix A is defined by
$$\|A\|_1 = \operatorname{tr}|A| = \operatorname{tr}\sqrt{A^* A}.$$
Defining $A := \rho - \sigma$, since A is self-adjoint, we may write its eigendecomposition,
$$A = \sum_i a_i |\phi_i\rangle\langle\phi_i| = \underbrace{\sum_{i:a_i>0} a_i |\phi_i\rangle\langle\phi_i|}_{Q} - \underbrace{\sum_{i:a_i\le 0} (-a_i) |\phi_i\rangle\langle\phi_i|}_{R}$$
where we have defined two positive semi-definite operators Q and R. Thus, $A = Q - R$ for $Q, R \ge 0$. This is sometimes called the Jordan decomposition of A. By using that the eigenvalues of the absolute value of A are the absolute value of the eigenvalues of A, we have
$$\operatorname{tr}|A| = \sum_i |a_i| = \operatorname{tr}(Q) + \operatorname{tr}(R)$$

and thus
$$D(\rho,\sigma) = \frac{1}{2}(\operatorname{tr} Q + \operatorname{tr} R).$$

2. Since $\operatorname{tr}(\rho - \sigma) = \operatorname{tr}(A) = 0$, we have $\operatorname{tr}(R - Q) = 0$ so $\operatorname{tr}(R) = \operatorname{tr}(Q)$, using the definitions from the previous part. Thus, $D(\rho, \sigma) = \operatorname{tr} Q$. First, if we choose P to be the projection onto the support of Q, namely $P = \sum_{i:a_i>0} |\phi_i\rangle\langle\phi_i|$, we have
$$\operatorname{tr}(P(\rho - \sigma)) = \operatorname{tr}[P(Q - R)] = \operatorname{tr}[PQ] - \operatorname{tr}[PR].$$
But since P and R are orthogonal by construction, $\operatorname{tr}[PR] = 0$. Moreover, since P is the projection onto the support of Q, $PQ = Q$ so $\operatorname{tr}[PQ] = \operatorname{tr}[Q] = D(\rho,\sigma)$. It remains to show for any projection P we have $\operatorname{tr}[P(\rho - \sigma)] \le D(\rho, \sigma)$. We have
$$\operatorname{tr}[P(\rho - \sigma)] = \operatorname{tr}[PQ] - \operatorname{tr}[PR] = \operatorname{tr}[PQ] - \operatorname{tr}[PRP] \le \operatorname{tr}[PQ]$$
since $P = P^2$ and using cyclicity of the trace, and the very useful property that if $A \ge B$ then $CAC^\dagger \ge CBC^\dagger$. Note that this is simply the positivity[1] (and linearity) of the map with one Kraus operator given by C. In particular, $R \ge 0$ means that $PRP^\dagger \ge 0$. Since $P = P^\dagger$, this shows $\operatorname{tr}[PRP] \ge 0$, yielding the inequality.

Next, since $P \le I$, we have $Q^{1/2} P Q^{1/2} \le Q$, so $\operatorname{tr}[PQ] = \operatorname{tr}[Q^{1/2} P Q^{1/2}] \le \operatorname{tr}[Q] = D(\rho, \sigma)$. Note we used that $Q \ge 0$ so that $Q^{1/2} \ge 0$ and in particular is Hermitian.

3. This is very similar to the previous step. Again, taking T as the projection onto the support of Q shows that we can obtain equality. The only step before where we used that $P = P^2$ is to show that $\operatorname{tr}[PR] \ge 0$. But instead of squaring P and cycling it, we can take the square root: $\operatorname{tr}[TR] = \operatorname{tr}[T^{1/2} R T^{1/2}] \ge 0$ using the same argument as before.

4. Let $|\psi^\rho_{RA}\rangle$ be a purification of ρ. Then the entanglement fidelity is defined as
$$F_e(\rho, \Lambda) := \langle\psi^\rho_{RA}|(\mathrm{id} \otimes \Lambda)\psi^\rho_{RA}|\psi^\rho_{RA}\rangle.$$
That is, $F_e(\rho, \Lambda) = F(\psi^\rho_{RA}, \sigma_{RA})^2$ for $\sigma = (\mathrm{id} \otimes \Lambda)\psi^\rho_{RA}$, where F is the fidelity. By the monotonicity of the fidelity under partial trace,
$$F(\psi^\rho_{RA}, \sigma_{RA}) \le F(\rho, \sigma_A).$$
Since $\sigma_A = \operatorname{tr}_R \sigma_{RA} = \Lambda(\rho)$, we obtain the result.

Problem 2.

1. State the Holevo-Schumacher-Westmoreland theorem.
   a. Use it to obtain the product-state classical capacity of a qubit depolarizing channel Λ defined as follows:
   $$\Lambda(\rho) = p\rho + \frac{1-p}{3}(\sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z),$$
   where σ_x, σ_y, and σ_z are the Pauli matrices.
   b. Consider an ensemble of quantum states $\mathcal{E} = \{p_x, \rho_x\}$ and let $\chi(\mathcal{E})$ denote its Holevo quantity. Let Λ be a quantum channel. Prove that
   $$\chi(\mathcal{E}') \le \chi(\mathcal{E})$$
   where $\mathcal{E}' = \{p_x, \Lambda(\rho_x)\}$.

[1] this map is in fact completely positive, since it has a Kraus decomposition; we only need positivity, however.

2. Consider a memoryless quantum information source characterized by $\{\pi, \mathcal{H}\}$, where $\pi \in \mathcal{D}(\mathcal{H})$. Suppose on n uses, the source emits a signal state $|\Psi_k^{(n)}\rangle \in \mathcal{H}^{\otimes n}$ with probability $p_k^{(n)}$, the index k labelling the different possible signal states. State the Typical Subspace Theorem, and use it prove that for such a source there exists a reliable compression-decompression scheme of rate $R > S(\pi)$ where $S(\pi)$ denotes the von Neumann entropy of the source.

Note: we won't cover part (2) in the revision class, since it's just bookwork.

Solution. 1. The HSW theorem says that the *product state classical capacity* of a quantum channel Λ is given by

$$C^{(1)}(\Lambda) = \chi^*(\Lambda)$$

where $\chi^*(\Lambda)$ is the Holevo capacity, defined as

$$\chi^*(\Lambda) := \max_{\{p_x,\rho_x\}} \chi(\{p_x, \Lambda(\rho_x)\}) \tag{1}$$

where $\chi(\cdot)$ is the Holevo χ-quantity, defined by

$$\chi(\{p_x, \Lambda(\rho_x)\}) := S(\Lambda(\sum_x p_x\rho_x)) - \sum_x p_x S(\Lambda(\rho_x))$$

and the maximum in (1) is taken over all ensembes $\{p_x, \rho_x\}$ of possible input states ρ_x of the channel, and $p_x \geq 0$, $\sum_x p_x = 1$.

a. Since

$$\frac{I}{2} = \frac{1}{4}[\rho + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z]$$

we find that

$$\begin{aligned}\Lambda(\rho) &= p\rho + \frac{1-p}{3}(2I - \rho)\\ &= \rho[p - \frac{1-p}{3}] + 4\left(\frac{1-p}{3}\right)\frac{I}{2}\\ &= q\rho + (1-q)\frac{I}{2}\end{aligned}$$

where $q = p - \frac{1-p}{3} = \frac{4p-1}{3}$. Let $\mathcal{E} = \{p_j, |\psi_j\rangle\langle\psi_j|\}$ be an input ensemble of pure states. Then

$$\Lambda(|\psi_j\rangle\langle\psi_j|) = q|\psi_j\rangle\langle\psi_j| + (1-q)\frac{I}{2}$$

which has eigenvalues $\frac{1\pm q}{2}$, regardless of which pure state was input. Thus, $S(\Lambda(|\psi_j\rangle\langle\psi_j|)) = h(\frac{1+q}{2})$ where $h(\cdot)$ denotes the binary Shannon entropy. Thus, since we can restrict the maximum in (1) to be only over pure states, we have

$$\begin{aligned}C^{(1)}(\Lambda) &= \chi^*(\Lambda)\\ &= \max_{\{p_j,\rho_j\}} S(\sum_j p_j\rho_j) - \sum_j p_j h(\frac{1+q}{2})\\ &= \max_{\{p_j,\rho_j\}} S(\sum_j p_j\rho_j) - h(\frac{1+q}{2})\\ &\leq \log 2 - h(\frac{1+q}{2})\\ &= 1 - h(\frac{1+q}{2})\end{aligned}$$

since the von Neummon entropy of a qubit is bounded by $\log 2$. By choosing the ensemble to be $p_1 = p_2 = \frac{1}{2}$ and $|\psi_1\rangle = |0\rangle$, $|\psi_2\rangle = |1\rangle$, we have $\sum_j p_j\rho_j = \frac{I}{2}$ which has von Neummon entropy $\log 2 = 1$. Since this upper bound is achievable, we thus have

$$C^{(1)}(\Lambda) = 1 - h(\frac{1+q}{2})$$

where $q = \frac{4p-1}{3}$.

b. We have that

$$\begin{aligned}\chi(\mathcal{E}) &= S(\sum_x p_x\rho_x) - \sum_x p_x S(\rho_x)\\ &= \sum_x p_x S(\rho_x\|\rho)\end{aligned}$$

where $\rho = \sum_x p_x\rho_x$. We can verify this by expanding the second form:

$$\begin{aligned}\sum_x p_x S(\rho_x\|\rho) &= \sum_x p_x \operatorname{tr}[\rho_x(\log\rho_x - \log\rho)]\\ &= \sum_x p_x(-S(\rho_x) - \operatorname{tr}[\rho_x\log\rho]\\ &= -\sum_x p_x S(\rho_x) - \sum_x p_x \operatorname{tr}[\rho_x\log\rho]\\ &= -\sum_x p_x S(\rho_x) - \operatorname{tr}[\rho\log\rho]\\ &= S(\rho) - \sum_x p_x S(\rho_x)\end{aligned}$$

as desired. Likewise, $\chi(\mathcal{E}') = \sum_x p_x S(\Lambda(\rho_x)\|\Lambda(\rho))$. But by the monotonicity of the relative entropy under CPTP maps (i.e. data-processing), we have

$$S(\Lambda(\rho_x)\|\Lambda(\rho)) \leq S(\rho_x\|\rho)$$

for each x. Multiplying by p_x and summing over x yields the result that $\chi(\mathcal{E}') \leq \chi(\mathcal{E})$.

2. This is the first half of Theorem 2 in Notes 14.

Problem 3.

1. The Bell states $|\Phi^+_{AB}\rangle$, $|\Phi^-_{AB}\rangle$, $|\Psi^+_{AB}\rangle$, $|\Psi^-_{AB}\rangle$, can be characterized by two classical bits, namely the parity bit and the phase bit. Show that the latter are eigenvalues of two commuting observables.
2. The Bell states form an orthonormal basis of the two-qubit Hilbert space. It is referred to as the Bell basis. Let us denote it by B_1. A sequence of two operations can be used to convert states of the computational basis $B_2 := \{|ij\rangle : i, j \in \{0, 1\}\}$ to the Bell states. State what these operations are. Can they also be used to convert states of B_1 to B_2? Justify your answer.
3. Prove that the Schmidt rank of a pure state *cannot* be increased by local operations and classical communication (LOCC), clearly stating any theorem that you use.
4. It is known that a matrix A is doubly stochastic if and only if $\vec{x} \prec \vec{y}$ for all vectors $\vec{y}$, where $x = A\vec{y}$.
   Let $\rho \in \mathcal{D}(\mathcal{H})$ be a state, where $\dim\mathcal{H} = d$, and let $\Lambda : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H})$ be a *unital* channel. Let $\vec{r} = (r_1, r_2, \ldots, r_d)$ and $\vec{s} = (s_1, s_2, \ldots, s_d)$ respectively denote the vectors of eigenvalues of ρ and $\sigma = \Lambda(\rho)$, arranged in non-increasing order. Using the above result, prove that $\vec{s} \prec \vec{r}$.
   Note: the question originally said $\vec{r} \prec \vec{s}$, which was an error.

Solution. 1. The parity bit is 0 if it is in a $|\Phi^{\pm}_{AB}\rangle$ state, and 1 if its in a $|\Psi^{\pm}_{AB}\rangle$ state. The phase bit is 0 if its in a $|\alpha^{+}\rangle$ state (for $\alpha \in \{\Phi, \Psi\}$), and 1 if it's in a $|\alpha^{-}\rangle$ state. Let $X_{AB} = \sigma_x^{(A)} \otimes \sigma_x^{(B)}$ where σ_x is the Pauli x matrix, and likewise $Z_{AB} = \sigma_z^{(A)} \otimes \sigma_z^{(B)}$. Then one can compute $[X_{AB}, Z_{AB}] = 0$[2]. Moreover,

$$X_{AB}|\alpha^{+}_{AB}\rangle = |\alpha^{+}_{AB}\rangle$$
$$X_{AB}|\alpha^{-}_{AB}\rangle = -|\alpha^{-}_{AB}\rangle$$

and thus the eigenvalues of X_{AB} are the phase bit. Likewise,

$$Z_{AB}|\Phi^{\pm}_{AB}\rangle = |\Phi^{\pm}_{AB}\rangle$$
$$Z_{AB}|\Psi^{\pm}_{AB}\rangle = -|\Psi^{\pm}_{AB}\rangle$$

so the eigenvalues of Z_{AB} are the parity bit.

Alternatively, one can consider the projections

$$P_{\text{parity}} = |\Psi^{+}_{AB}\rangle\langle\Psi^{+}_{AB}| + |\Psi^{-}_{AB}\rangle\langle\Psi^{-}_{AB}|$$

and

$$P_{\text{phase}} = |\Phi^{-}_{AB}\rangle\langle\Phi^{-}_{AB}| + |\Psi^{-}_{AB}\rangle\langle\Psi^{-}_{AB}|.$$

Then the eigenvalues of these operators are exactly the phase and parity bit, in the sense that

$$P_{\text{parity}}|\Psi^{\pm}_{AB}\rangle = |\Psi^{\pm}_{AB}\rangle, \qquad P_{\text{parity}}|\Phi^{\pm}_{AB}\rangle = 0,$$

and for $\alpha \in \{\Phi, \Psi\}$,

$$P_{\text{phase}}|\alpha^{-}\rangle = |\alpha^{-}\rangle, \qquad P_{\text{phase}}|\alpha^{+}\rangle = 0.$$

Moreover these projections commute.

2. Hadamard then CNOT. It is reversible.

3. Nielsen's Majorization Theorem states that a bipartite state $|\psi_{AB}\rangle$ can be converted to $|\phi_{AB}\rangle$ if and only if

$$\lambda_\psi \prec \lambda_\phi$$

where λ_ψ and λ_ϕ are the vectors of eigenvalues of $\psi_A := \operatorname{tr}_B |\psi_{AB}\rangle\langle\psi_{AB}|$ and $\phi_A = \operatorname{tr}_B |\phi_{AB}\rangle\langle\phi_{AB}|$ and $\prec$ is the majorization pre-order. Let $n(\psi)$ and $n(\phi)$ denote the Schmidt ranks of two pure states $|\psi_{AB}\rangle$ and $|\phi\rangle_{AB}$, such that $|\psi\rangle_{AB} \xrightarrow{\text{LOCC}} |\phi_{AB}\rangle$. Assume

$$n(\psi) < n(\phi). \tag{2}$$

Let $\lambda_\psi = (\nu_1, \ldots, \nu_d)$ and $\lambda_\phi = (\mu_1, \ldots, \mu_d)$, where $d = \dim \mathcal{H}_A$ is the dimension of the Hilbert space for system A. The assumption (2) implies that there exists some integer $m \leq d$ such that $\mu_m \neq 0$ but $\nu_m = 0$. Hence

$$\sum_{i=1}^{m-1} \nu_i = 1, \qquad \text{but} \qquad \sum_{i=1}^{m-1} \mu_i < 1.$$

This contradicts that $\lambda_\psi \prec \lambda_\phi$. Therefore, (2) cannot hold, and therefore the Schmidt rank cannot increase by LOCC operations.

[2] Takes a little time and expansion; this question was from an earlier iteration of the class where this was treated.

4. Let us write the eigendecompositions

$$\rho = \sum_{i=1}^{d} r_i |e_i\rangle\langle e_i|$$
$$\sigma = \sum_{i=1}^{d} s_j |f_j\rangle\langle f_j|$$

where the eigenvalues are arranged in non-increasing order. We have that

$$\begin{aligned} s_j &= \operatorname{tr}[\sigma |f_j\rangle\langle f_j|] \\ &= \operatorname{tr}[\Lambda(\rho)|f_j\rangle\langle f_j|] \quad = \sum_i r_i \operatorname{tr}[\Lambda(|e_i\rangle\langle e_i|)|f_j\rangle\langle f_j|]. \end{aligned}$$

Now define a matrix D with entries $D_{ji} = \operatorname{tr}[\Lambda(|e_i\rangle\langle e_i|)|f_j\rangle\langle f_j|]$. Then we've shown

$$s_j = \sum_i D_{ji} r_i.$$

Thus, it remains to show that D is doubly-stochastic, by the result quoted in the question. We have

$$\begin{aligned} \sum_j D_{ji} &= \sum_j \operatorname{tr}[\Lambda(|e_i\rangle\langle e_i|)|f_j\rangle\langle f_j|] \\ &= \operatorname{tr}[\Lambda(|e_i\rangle\langle e_i|)I] = \operatorname{tr}[\Lambda(|e_i\rangle\langle e_i|)] = 1 \end{aligned}$$

since Λ is trace-preserving. Next,

$$\begin{aligned} \sum_i D_{ji} &= \sum_i \operatorname{tr}[\Lambda(|e_i\rangle\langle e_i|)|f_j\rangle\langle f_j|] \\ &= \operatorname{tr}[\Lambda(I)|f_j\rangle\langle f_j|] = \operatorname{tr}[|f_j\rangle\langle f_j|] = 1 \end{aligned}$$

since Λ is unital. Therefore, D is doubly stochastic. Since $\vec{s} = D\vec{r}$, we have $\vec{s} \prec \vec{r}$.

Problem 4.

1. Prove that any completely positive trace-preserving map Φ, acting on states ρ in a Hilbert space $\mathcal{H}_A$ can be written in the Kraus form:

$$\Phi(\rho) = \sum_k A_k \rho A_k^\dagger,$$

where A_k are linear operators satisfying $\sum_k A_k^\dagger A_k = I$ and I is the identity operator.

Hint: consider a maximally entangled state.

2. Using a maximally entangled state and the properties of the swap operator, prove that the transposition operator T is positive but not completely positive.

Solution. 1. See Theorem 1 of Notes 9.

2. First, the transpose operator on square matrices is positive because it is spectrum preserving. We can see this by noticing for a matrix A, the eigenvalues of A are the roots of its characteristic polynomial $\lambda \mapsto \det(\lambda I - A)$. But

$$\det(\lambda I - A^T) = \det((\lambda I - A)^T) = \det(\lambda I - A)$$

since transpose is linear and determinant is invariant under transpose. Thus, both A and its transpose have the same characeristic polynomial and therefore the same eigenvalues. Hence the transpose maps positive semi-definite matrices to positive semi-definite matrices.

Next, the swap operator acts as $\mathbb{F}|i\rangle \otimes |j\rangle = |j\rangle \otimes |i\rangle$. We can then write a matrix representation of the swap operator as

$$\mathbb{F} = \sum_{i,j} |ji\rangle\langle ij| = \sum_{i,j} |ij\rangle\langle ji|$$

where we have swiched the roles of the indices i and j in the second formulation. By comparing to the MES in this basis, $\Omega = \frac{1}{d}\sum_{i,j} |ij\rangle\langle ij|$, we see $\frac{1}{d}\mathbb{F} = (\mathrm{id} \otimes T)\Omega$, where T is the transpose operator in the basis $\{|i\rangle\}$. If T were completely positive, then since $\Omega \geq 0$, we would have $(\mathrm{id} \otimes T)\Omega \geq 0$. But this cannot be since $(\mathrm{id} \otimes T) = \frac{1}{d}\mathbb{F}$ and the swap operator has negative eigenvalues. We can see this by noticing that $\mathbb{F}$ is self-adjoint so it has real eiegnvalues and squares to the identity, so all its eigenvalues must square to 1. Therefore, all of its eigenvalues must be ± 1. But if all its eigenvalues were $+1$, then $\mathbb{F} = I$ which is not the case; thus, $\mathbb{F}$ has negative eigenvalues, so T is not completely positive.

Problem 5.

Consider the decay of a two-level atom from its excited state to its ground state. Let the probability of this decay be p. The spontaneous emission of a photon accompanies this decay.

1. Name the quantum channel that can be used to model this process and write its Kraus operators. What process does each of these Kraus operators correspond to? Give reasons for your answer.
2. What is a unital channel? Is the channel in (1) unital?
3. Suppose the atom is originally in a state $\rho := \sum_{\alpha,\beta=0}^{1} \rho_{\alpha\beta}|\alpha\rangle\langle\beta|$, where $|\alpha\rangle, |\beta\rangle$, $\alpha, \beta \in \{0, 1\}$, denote orthonormal basis states of the Hilbert space of the atom. By considering the action of a unitary operator U on the atom and its environment, deduce how the state of the atom changes under the action of one use of the channel in (1).
4. Let AB denote a bipartite system which is in state ρ_{AB}. Show that the mutual information of the system cannot increase under the action of a completely positive trace-preserving map on the subsystem B alone.

Solution.

1. This is called the amplitude damping channel. It has two Kraus operators,

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}$$

in the basis where $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ corresponds to the ground state of the atom, and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ corresponds to the excited state.
The Kraus operator A_1 corresponds to the process that the ground state of the system remains invariant while the excited state is invariant only with probability $1 - p$, since $A_1|0\rangle = |0\rangle$ while $A_1|1\rangle = \sqrt{1-p}|1\rangle$. The operator A_2 corresponds to the decay of the excited state to the ground state with probability p, since $A_2|0\rangle = 0$ while $A_2|1\rangle = \sqrt{p}|0\rangle$.

2. A unital channel is a channel Λ such that $\Lambda(I) = I$, i.e. it preserves the identity matrix. The amplitude damping channel is not unital, since

$$A_1 I A_1^\dagger + A_2 I A_2^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & 1-p \end{pmatrix} + \begin{pmatrix} p & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1+p & 0 \\ 0 & 1-p \end{pmatrix}$$

which is not the identity matrix for $p \neq 0$.

3. We can write

$$\rho_A = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}$$

where we write A to label the Hilbert space of the atom. Let us write E for the environment and choose a bipartite unitary operator U_{AE} such that $U_{AE}|00\rangle_{AE} = |00\rangle_{AE}$ and $U_{AE}|10\rangle_{AE} = \sqrt{1-p}|10\rangle_{AE} + \sqrt{p}|01\rangle_{AE}$, where $|0\rangle_E$ is the vacuum state of the photon, and $|1\rangle_E$ the excited state corresponding to the emission of the photon.

As an aside, it was pointed out in the revision class that this definition does not completely specify the unitary; we only say what happens when the environment is in state $|0\rangle_E$, since that's all we need to model the interaction between the atom (in any state) and the environment when the environment starts in the state $|0\rangle_E$. To truly construct a unitary, however, one needs to specify what happens when the photon is in excited state. I believe there may be more than one choice that gives a unitary evolution; one choice[3] is

$$U|11\rangle = |11\rangle, \qquad U|01\rangle = \sqrt{1-p}|01\rangle - \sqrt{p}|10\rangle.$$

Getting back to the problem, we consider the joint final state,

$$\sigma_{AE} := U_{AE}\rho_A \otimes |0\rangle\langle 0|_E U_{AE}^\dagger$$

which models the state of the atom and the photon after the interaction. We

$$\begin{aligned}
\sigma_{AE} &= \sum_{\alpha,\beta=0}^{1} U_{AE}\rho_{\alpha\beta}|\alpha\rangle\langle\beta|_A \otimes |0\rangle\langle 0|_E U_{AE}^\dagger \\
&= \rho_{00}U_{AE}|00\rangle\langle 00|U_{AE}^\dagger + \rho_{10}U_{AE}|10\rangle\langle 00|U_{AE}^\dagger + \rho_{01}U_{AE}|00\rangle\langle 10|U_{AE}^\dagger + \rho_{11}U_{AE}|10\rangle\langle 10|U_{AE}^\dagger \\
&= \rho_{00}|00\rangle\langle 00| + \rho_{10}(\sqrt{1-p}|10\rangle + \sqrt{p}|01\rangle)\langle 00| \\
&\quad + \rho_{01}|00\rangle(\sqrt{1-p}\langle 10| + \sqrt{p}\langle 01|) + \rho_{11}(\sqrt{1-p}|10\rangle + \sqrt{p}|01\rangle)(\sqrt{1-p}\langle 10| + \sqrt{p}\langle 01|)
\end{aligned}$$

where we have dropped the labels A and E for ease of notation. Now, we can trace out E term by term to find

$$\begin{aligned}
\mathrm{tr}_E\, \sigma_{AE} &= \rho_{00}|0\rangle\langle 0| + \rho_{10}\sqrt{1-p}|1\rangle\langle 0| + \rho_{01}\sqrt{1-p}|0\rangle\langle 1| + \rho_{11}[p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|] \\
&= \begin{pmatrix} \rho_{00} + p\rho_{11} & \sqrt{1-p}\rho_{10} \\ \sqrt{1-p}\rho_{01} & (1-p)\rho_{11} \end{pmatrix}
\end{aligned}$$

which describes the result of applying the channel once to the input state ρ.

4. We can write the mutual information of a state ρ_{AB} in terms of the relative entropy, as

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

Likewise, if we apply a quantum channel Λ to the be system alone, we apply $\mathrm{id} \otimes \Lambda$ to the bipartite state ρ_{AB}, yielding $\sigma_{AB} := \mathrm{id} \otimes \Lambda(\rho_{AB})$. So the mutual information can be written

$$I(A:B)_\sigma = D(\sigma_{AB}\|\sigma_A \otimes \sigma_B).$$

But $\sigma_A = \rho_A$, since only the identity channel was applied on the A part, while $\sigma_B = \Lambda(\rho_B)$. Thus,

$$\begin{aligned}
I(A:B)_\sigma &= D(\sigma_{AB}\|\rho_A \otimes \Lambda(\rho_B)) \\
&= D(\mathrm{id} \otimes \Lambda(\rho_{AB})\|\,\mathrm{id} \otimes \Lambda(\rho_A \otimes \rho_B)).
\end{aligned}$$

[3]This choice seems fairly obvious in retrospect, but I found it by guessing and checking with my current favorite scientific programming language, Julia.

By the data processing inequality for the relative entropy, since $\mathrm{id} \otimes \Lambda$ is a CPTP map, we have

$$D(\mathrm{id} \otimes \Lambda(\rho_{AB}) \| \mathrm{id} \otimes \Lambda(\rho_A \otimes \rho_B)) \leq D(\rho_{AB} \| \rho_A \otimes \rho_B)$$

which proves $I(A:B)_\sigma \leq I(A:B)_\rho$ as desired.