University of Oslo

FYS3415/FYS4415 — Quantum computing and quantum information

Problem set 12

This problem set will be discussed on **May 6**.

Unless otherwise stated, all logarithms are base 2, so entropies are measured in bits.

Problem 1: Quantum Relative Entropy and Mutual Information

In the previous set you studied the von Neumann entropy and proved a number of structural results, such as the joint entropy theorem and the equality of entropies for purifications. In this problem you continue along these lines and develop the *quantum relative entropy*, which plays the same fundamental role in quantum information that the classical relative entropy plays in Shannon theory.

For two density operators ρ and σ on the same Hilbert space, the quantum relative entropy is

$$S(\rho \,\|\, \sigma) \equiv \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(\rho \log \sigma),$$

defined to be $+\infty$ when the support of ρ is not contained in the support of σ. Just as in the classical case, $S(\rho \,\|\, \sigma) \geq 0$ with equality iff $\rho = \sigma$ (*Klein's inequality*); we will take this as a given here.

a) Relative entropy in the commuting case.

Suppose ρ and σ commute, so they share an orthonormal eigenbasis $\{|x\rangle\}$ with eigenvalues $p(x)$ and $q(x)$, respectively.

Use the usual conventions $0 \log(0/q) = 0$, and if $p(x) > 0$ while $q(x) = 0$, the corresponding contribution is $+\infty$.

i) Show that

$$S(\rho \,\|\, \sigma) = H\big(p(x) \,\|\, q(x)\big),$$

where the right-hand side is the classical relative entropy of the eigenvalue distributions.

ii) Conclude that $S(\rho \,\|\, \sigma) \geq 0$ in this case, and identify when equality holds.

iii) As a concrete check, compute $S(\rho \,\|\, \sigma)$ for the qubit pair $\rho = |0\rangle\langle 0|$ and $\sigma = I/2$, and verify that the result is non-negative.

Solution: If some $p(x) > 0$ while $q(x) = 0$, both the quantum and classical relative entropies are $+\infty$. Otherwise, we may compute directly.

i) Since ρ and σ are simultaneously diagonal, $\log \rho = \sum_x \log p(x)\ |x\rangle\langle x|$ and similarly for $\log \sigma$. Therefore

$$\begin{aligned} S(\rho \,\|\, \sigma) &= \operatorname{Tr}\left(\sum_x p(x) \log p(x)\ |x\rangle\langle x| \right) - \operatorname{Tr}\left(\sum_x p(x) \log q(x)\ |x\rangle\langle x| \right) \\ &= \sum_x p(x) \log p(x) - \sum_x p(x) \log q(x) = \sum_x p(x) \log \frac{p(x)}{q(x)} = H\big(p(x) \,\|\, q(x)\big). \end{aligned}$$

ii) Non-negativity is the classical relative entropy inequality proven in problem set 11. Equality holds iff $p(x) = q(x)$ for all x, i.e. iff $\rho = \sigma$.

iii) Both states are diagonal in the computational basis with eigenvalues $p(0) = 1$, $p(1) = 0$ and $q(0) = q(1) = 1/2$, so they commute and part i) applies. With the convention $0 \log 0 = 0$,

$$S(\rho \,\|\, \sigma) = 1 \cdot \log \frac{1}{1/2} + 0 \cdot \log \frac{0}{1/2} = \log 2 = 1 \text{ bit}.$$

This is positive, as expected, and equals the entropy difference between the pure state and the maximally mixed qubit.

b) Quantum mutual information and subadditivity.

For a bipartite state ρ^{AB} with reduced states $\rho^A = \mathrm{Tr}_B\, \rho^{AB}$ and $\rho^B = \mathrm{Tr}_A\, \rho^{AB}$, the *quantum mutual information* is

$$I(A:B) \equiv S(A) + S(B) - S(A,B).$$

i) Show that $I(A:B) = S\big(\rho^{AB} \,\big\|\, \rho^A \otimes \rho^B\big)$.

ii) Conclude the *subadditivity* inequality $S(A,B) \leq S(A) + S(B)$, and identify when equality holds.

Hint[1].

Solution: i) Write

$$S(\rho^{AB} \,\|\, \rho^A \otimes \rho^B) = \mathrm{Tr}\big(\rho^{AB} \log \rho^{AB}\big) - \mathrm{Tr}\big(\rho^{AB} \log\big(\rho^A \otimes \rho^B\big)\big).$$

The first term is $-S(A,B)$. For the second, the hint gives

$$\log\big(\rho^A \otimes \rho^B\big) = \log \rho^A \otimes I + I \otimes \log \rho^B,$$

so

$$\begin{aligned} \mathrm{Tr}\big(\rho^{AB} \log\big(\rho^A \otimes \rho^B\big)\big) &= \mathrm{Tr}\big(\rho^{AB}(\log \rho^A \otimes I)\big) + \mathrm{Tr}\big(\rho^{AB}(I \otimes \log \rho^B)\big) \\ &= \mathrm{Tr}\big(\rho^A \log \rho^A\big) + \mathrm{Tr}\big(\rho^B \log \rho^B\big) = -S(A) - S(B), \end{aligned}$$

where in the second line we used

$$\mathrm{Tr}_{AB}\big[\rho^{AB}(M \otimes I)\big] = \mathrm{Tr}_A(\rho^A M), \qquad \mathrm{Tr}_{AB}\big[\rho^{AB}(I \otimes N)\big] = \mathrm{Tr}_B(\rho^B N).$$

Therefore

$$S(\rho^{AB} \,\|\, \rho^A \otimes \rho^B) = -S(A,B) + S(A) + S(B) = I(A:B).$$

ii) By Klein's inequality the relative entropy is non-negative, so $I(A:B) \geq 0$ and hence

$$S(A,B) \leq S(A) + S(B).$$

Equality holds iff $\rho^{AB} = \rho^A \otimes \rho^B$, i.e. iff the two subsystems are uncorrelated.

c) A bipartite computation.

Consider the bipartite state

$$\rho^{AB} \;=\; \tfrac{1}{2}\, |0\rangle\langle 0| \otimes |0\rangle\langle 0| \;+\; \tfrac{1}{2}\, |1\rangle\langle 1| \otimes |+\rangle\langle +|,$$

[1] Use $\log\big(\rho^A \otimes \rho^B\big) = \log \rho^A \otimes I + I \otimes \log \rho^B$, which follows from the spectral decomposition of $\rho^A \otimes \rho^B$. The logarithms are understood on the supports of the corresponding density operators.

which can be read as Alice flipping a fair coin, recording the result on system A, and preparing system B in $|0\rangle$ if the coin came up 0 and $|+\rangle$ if it came up 1.

i) Compute the reduced states ρ^A and ρ^B and their von Neumann entropies $S(A)$ and $S(B)$. (You may reuse the eigenvalues of $\bar{\rho} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +|$ from problem set 11, problem 2a.)

ii) Find the eigenvalues of ρ^{AB} and use them to compute $S(A, B)$.

iii) Combine your results to evaluate $I(A:B)$. Compare to $S(A) = 1$ bit and explain in one sentence why the two are different.

Hint[2].

Solution: i) Tracing out B,

$$\rho^A = \tfrac{1}{2}|0\rangle\langle 0| + \tfrac{1}{2}|1\rangle\langle 1| = \tfrac{I}{2}, \qquad S(A) = \log 2 = 1 \text{ bit}.$$

Tracing out A,

$$\rho^B = \tfrac{1}{2}|0\rangle\langle 0| + \tfrac{1}{2}|+\rangle\langle +|,$$

which is exactly the state $\bar{\rho}$ from problem set 11, problem 2a. Its eigenvalues are $\lambda_\pm = (1 \pm 1/\sqrt{2})/2$, so

$$S(B) = h_{\mathrm{b}}\left(\frac{1 + 1/\sqrt{2}}{2}\right) \approx 0.601 \text{ bits}.$$

ii) Define

$$|v_0\rangle = |0\rangle_A \otimes |0\rangle_B, \qquad |v_1\rangle = |1\rangle_A \otimes |+\rangle_B.$$

Then

$$\rho^{AB} = \frac{1}{2}|v_0\rangle\langle v_0| + \frac{1}{2}|v_1\rangle\langle v_1|.$$

The two states are orthogonal, since

$$\langle v_0|v_1\rangle = \langle 0|1\rangle_A \langle 0|+\rangle_B = 0.$$

Thus this is already the spectral decomposition on its support. The non-zero eigenvalues are $1/2$ and $1/2$, so

$$S(A, B) = -2\left(\frac{1}{2}\log\frac{1}{2}\right) = \log 2 = 1 \text{ bit}.$$

iii)

$$I(A:B) = S(A) + S(B) - S(A, B) = 1 + 0.601 - 1 \approx 0.601 \text{ bits}.$$

Although A stores one full classical bit about which state was prepared, the register B is correlated with that label by only ≈ 0.601 bits in the Holevo/mutual-information sense, because the two possible preparations $|0\rangle$ and $|+\rangle$ have non-zero overlap $\langle 0|+\rangle = 1/\sqrt{2}$ and so are not perfectly distinguishable. We will recognize this number again in the next problem under the name *Holevo information*.

[2] For ii), notice that ρ^{AB} is a sum of two rank-one terms whose supports lie in mutually orthogonal subspaces of $\mathcal{H}_A \otimes \mathcal{H}_B$ (because $\langle 0|1\rangle_A = 0$). The eigenvalues of such a sum are simply the weights of the rank-one pieces.

Problem 2: Holevo Information and the Holevo Bound

Suppose Alice prepares a quantum state ρ_x with probability p_x, drawn from an ensemble $\mathcal{E} = \{p_x, \rho_x\}$, and sends it to Bob. Bob performs a POVM $\{E_y\}$ and obtains classical outcome Y. How much classical information about Alice's choice X can Bob's measurement reveal? The *Holevo bound* answers this question:

$$I(X:Y) \leq \chi(\mathcal{E}), \qquad \chi(\mathcal{E}) \equiv S\left(\sum_x p_x \rho_x\right) - \sum_x p_x\, S(\rho_x).$$

The quantity $\chi(\mathcal{E})$ is the *Holevo information* of the ensemble; the inequality holds for every choice of POVM.

a) Holevo information as a mutual information.

Introduce a "label" register A with orthonormal basis $\{|x\rangle\}$, and form the classical-quantum state

$$\rho^{AB} \;=\; \sum_x p_x\; |x\rangle\langle x| \otimes \rho_x$$

on $A \otimes B$. Compute ρ^A, ρ^B, and $S(A,B)$, using the joint entropy theorem

$$S(A,B) = H(p_x) + \sum_x p_x S(\rho_x).$$

Then show that

$$I(A:B) = \chi(\mathcal{E}).$$

Solution: Tracing out system B gives

$$\rho^A = \mathrm{Tr}_B(\rho^{AB}) = \sum_x p_x\, |x\rangle\langle x|\, \mathrm{Tr}(\rho_x) = \sum_x p_x\, |x\rangle\langle x|,$$

so

$$S(A) = H(p_x).$$

Tracing out system A gives

$$\rho^B = \mathrm{Tr}_A(\rho^{AB}) = \sum_x p_x \rho_x \equiv \bar{\rho},$$

so

$$S(B) = S(\bar{\rho}).$$

The joint entropy theorem gives

$$S(A,B) = H(p_x) + \sum_x p_x S(\rho_x).$$

Therefore

$$\begin{aligned} I(A:B) &= S(A) + S(B) - S(A,B) \\ &= H(p_x) + S(\bar{\rho}) - \left(H(p_x) + \sum_x p_x S(\rho_x)\right) \\ &= S(\bar{\rho}) - \sum_x p_x S(\rho_x) = \chi(\mathcal{E}). \end{aligned}$$

Thus the Holevo information is the quantum mutual information between the classical label register and the quantum system that Bob receives.

b) The ensemble from problem 1c.

The bipartite state in problem 1c arose from the ensemble

$$\mathcal{E} = \left\{ \left(\tfrac{1}{2}, |0\rangle\langle 0|\right), \left(\tfrac{1}{2}, |+\rangle\langle +|\right) \right\}.$$

Compute $\chi(\mathcal{E})$ from the formula in part a), and check that it equals the $I(A:B)$ computed there directly.

Solution: Both preparations are pure, so $\sum_i p_i\, S(\rho_i) = 0$, and

$$\chi(\mathcal{E}) = S(\bar{\rho}) = h_{\mathrm{b}}\left(\frac{1+1/\sqrt{2}}{2}\right) \approx 0.601 \text{ bits}.$$

This is exactly the value of $I(A:B)$ computed in problem 1c by direct diagonalization. Since $H(p_i) = 1$ bit, the strict inequality $\chi < H(p_i)$ reflects the non-zero overlap $\langle 0|+\rangle = 1/\sqrt{2}$: no measurement on one copy of B can reveal one full bit about Alice's label.

c) A three-state example.

Consider the three pure qubit states

$$|\psi_0\rangle = |0\rangle, \qquad |\psi_1\rangle = -\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle, \qquad |\psi_2\rangle = -\frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle,$$

each prepared with probability 1/3.

i) Compute the average state

$$\bar{\rho} = \frac{1}{3}\sum_{x=0}^{2} |\psi_x\rangle\langle\psi_x|$$

and the Holevo information χ of this ensemble.

ii) Now define states $|\phi_x\rangle$ orthogonal to $|\psi_x\rangle$:

$$|\phi_0\rangle = |1\rangle, \qquad |\phi_1\rangle = \frac{\sqrt{3}}{2}|0\rangle + \frac{1}{2}|1\rangle, \qquad |\phi_2\rangle = \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle.$$

Bob measures with the POVM

$$E_x = \frac{2}{3}|\phi_x\rangle\langle\phi_x|, \qquad x = 0, 1, 2.$$

You may use that $\sum_x E_x = I$. Show that

$$p(Y = y \mid X = x) = \begin{cases} 0, & y = x, \\ 1/2, & y \neq x, \end{cases}$$

and compute the resulting classical mutual information $I(X:Y)$. Compare it with χ.

Solution: i) We compute the average state. The projectors onto $|\psi_1\rangle$ and $|\psi_2\rangle$ are

$$|\psi_1\rangle\langle\psi_1| = \frac{1}{4}|0\rangle\langle 0| - \frac{\sqrt{3}}{4}|0\rangle\langle 1| - \frac{\sqrt{3}}{4}|1\rangle\langle 0| + \frac{3}{4}|1\rangle\langle 1|,$$

and

$$|\psi_2\rangle\langle\psi_2| = \frac{1}{4}|0\rangle\langle 0| + \frac{\sqrt{3}}{4}|0\rangle\langle 1| + \frac{\sqrt{3}}{4}|1\rangle\langle 0| + \frac{3}{4}|1\rangle\langle 1|.$$

Adding these to $|\psi_0\rangle\langle\psi_0| = |0\rangle\langle 0|$, the off-diagonal terms cancel and we get

$$\sum_{x=0}^{2}|\psi_x\rangle\langle\psi_x| = \frac{3}{2}|0\rangle\langle 0| + \frac{3}{2}|1\rangle\langle 1| = \frac{3}{2}I.$$

Thus

$$\bar{\rho} = \frac{I}{2}.$$

All three preparations are pure, so $S(|\psi_x\rangle\langle\psi_x|) = 0$ for all x. Therefore

$$\chi = S(\bar{\rho}) - \frac{1}{3}\sum_{x=0}^{2}S(|\psi_x\rangle\langle\psi_x|) = S(I/2) = 1 \text{ bit}.$$

ii) Since $|\phi_x\rangle$ is orthogonal to $|\psi_x\rangle$,

$$p(Y = x \mid X = x) = \langle\psi_x|E_x|\psi_x\rangle = \frac{2}{3}|\langle\phi_x|\psi_x\rangle|^2 = 0.$$

For $y \neq x$, one checks that

$$|\langle\phi_y|\psi_x\rangle|^2 = \frac{3}{4},$$

so

$$p(Y = y \mid X = x) = \langle\psi_x|E_y|\psi_x\rangle = \frac{2}{3}\cdot\frac{3}{4} = \frac{1}{2}.$$

Thus, conditioned on $X = x$, Bob's outcome is equally likely to be one of the two labels different from x. Hence

$$H(Y|X = x) = 1 \text{ bit}$$

for every x, and so

$$H(Y|X) = 1 \text{ bit}.$$

By symmetry, Y is uniformly distributed over three outcomes, so

$$H(Y) = \log 3.$$

Therefore

$$I(X:Y) = H(Y) - H(Y|X) = \log 3 - 1 \approx 0.585 \text{ bits}.$$

This is strictly smaller than the Holevo quantity:

$$I(X:Y) \approx 0.585 < 1 = \chi.$$

So the Holevo quantity is an upper bound on accessible classical information, not a guarantee that a measurement will extract χ bits. In fact, for this symmetric three-state ensemble the accessible information is $\log 3 - 1$, so the Holevo bound is not tight here; we will not prove optimality.

Problem 3: The Threshold Theorem and the Cost of Fault Tolerance

Real qubits decohere and real gates fail. A representative superconducting-transmon estimate might use a coherence time $T_2 \sim 100\,\mu$s and gate time $\tau_g \sim 25\,$ns, giving the crude coherence budget

$$T_2/\tau_g \sim \frac{100\,\mu\text{s}}{25\,\text{ns}} \approx 4 \times 10^3$$

gate times. Useful algorithms require vastly longer computations: a rough classroom estimate for Shor's algorithm on an n-bit integer scales like n^3 logical gates, so RSA-2048 suggests a scale around 10^{10} logical gates.

The *quantum threshold theorem* says that if the physical error rate is below a code-dependent threshold, then encoding and repeated error correction can suppress logical errors to arbitrarily low levels, at the price of overhead. In this problem we study this scaling using concatenated codes.

a) One level of encoding.

Recall that the 3-qubit bit-flip code corrects zero or one bit flip, but fails if two or three physical qubits flip. If each physical qubit flips independently with probability p, the logical error rate is

$$p_L = \binom{3}{2} p^2(1-p) + p^3 = 3p^2 - 2p^3.$$

i) Solve $p_L < p$. For which physical error rates does the encoding improve the qubit?

ii) For $p \ll 1$, we have $p_L \approx 3p^2$. Show that this can be written as

$$\frac{p_L}{p_{\text{th}}} \approx \left(\frac{p}{p_{\text{th}}}\right)^2, \qquad p_{\text{th}} = \frac{1}{3}.$$

Why is this p_{th} not the same as the exact crossover found in part i)?

Solution: i) The condition $p_L < p$ gives

$$3p^2 - 2p^3 < p,$$

or

$$p(1-p)(1-2p) > 0.$$

For $0 < p < 1$, this is equivalent to

$$p < \frac{1}{2}.$$

Thus the code improves the error rate only when the physical bit-flip probability is below $1/2$. At $p = 1/2$, it gives $p_L = p$, so it is not beneficial. For $1/2 < p < 1$, this encoding is harmful under the simple bit-flip noise model.

ii) If $p_L \approx 3p^2$ and $p_{\text{th}} = 1/3$, then

$$\frac{p_L}{p_{\text{th}}} = \frac{3p^2}{1/3} = 9p^2 = \left(\frac{p}{1/3}\right)^2 = \left(\frac{p}{p_{\text{th}}}\right)^2.$$

The value $p_{\text{th}} = 1/3$ is the threshold of the leading-order recursion $p_L \approx 3p^2$. It is not the exact one-level crossover, because the exact formula contains the cubic correction $-2p^3$, which shifts the crossover to $p = 1/2$.

b) Concatenation.

In concatenated coding, each physical qubit is itself replaced by an encoded block. Assume that after one level the effective error rate obeys

$$\frac{p^{(k)}}{p_{\text{th}}} = \left(\frac{p^{(k-1)}}{p_{\text{th}}}\right)^2, \qquad p^{(0)} = p.$$

i) Show by induction that

$$p^{(k)} = p_{\text{th}} \left(\frac{p}{p_{\text{th}}}\right)^{2^k}.$$

ii) Explain why the logical error rate decreases double-exponentially in k when $p < p_{\text{th}}$, and why concatenation fails to help when $p \geq p_{\text{th}}$.

Solution: i) The formula is true for $k = 0$. Assume

$$\frac{p^{(k-1)}}{p_{\text{th}}} = \left(\frac{p}{p_{\text{th}}}\right)^{2^{k-1}}.$$

Then

$$\frac{p^{(k)}}{p_{\text{th}}} = \left(\frac{p^{(k-1)}}{p_{\text{th}}}\right)^2 = \left[\left(\frac{p}{p_{\text{th}}}\right)^{2^{k-1}}\right]^2 = \left(\frac{p}{p_{\text{th}}}\right)^{2^k}.$$

Therefore

$$p^{(k)} = p_{\text{th}} \left(\frac{p}{p_{\text{th}}}\right)^{2^k}.$$

ii) If $p < p_{\text{th}}$, then $p/p_{\text{th}} < 1$, and each level squares this ratio. Since the exponent is 2^k, the suppression is double-exponential in the number of concatenation levels. If $p = p_{\text{th}}$, the recursion gives no suppression. If $p > p_{\text{th}}$, the ratio is larger than one and the same recursion no longer reduces errors.

c) A concatenated-code overhead estimate.

The 3-qubit bit-flip code is only a toy model: it corrects only X-errors and assumes perfect syndrome extraction. A simple concatenated-code benchmark is a $[[7, 1, 3]]$-style code, where each concatenation level multiplies the number of physical data qubits by 7. Thus a level-k logical qubit uses 7^k physical data qubits. Use the rough threshold

$$p_{\text{th}} = 10^{-4}.$$

i) A benchmark physical error rate is $p = 10^{-3}$. Is this below threshold for this concatenated-code model? What does that imply?

ii) Now suppose improved hardware reaches

$$p = 10^{-5}.$$

Using

$$p^{(k)} = p_{\text{th}} \left(\frac{p}{p_{\text{th}}}\right)^{2^k},$$

find the smallest k such that

$$p^{(k)} \leq 10^{-15}.$$

iii) For this k, estimate the number of physical data qubits needed for 10^4 logical qubits. Briefly state what this estimate ignores.

Hint[3].

Solution: i) Since

$$10^{-3} > 10^{-4},$$

the benchmark value $p = 10^{-3}$ is above threshold for this simplified concatenated-code model. The threshold theorem does not say that error correction always helps; it says that scalable suppression is possible once the physical error rate is below the relevant threshold.

ii) Now

$$\frac{p}{p_{\text{th}}} = \frac{10^{-5}}{10^{-4}} = 10^{-1}.$$

Therefore

$$p^{(k)} = 10^{-4}(10^{-1})^{2^k} = 10^{-4-2^k}.$$

We need

$$10^{-4-2^k} \leq 10^{-15},$$

so

$$4 + 2^k \geq 15, \qquad 2^k \geq 11.$$

The smallest such integer is

$$k = 4.$$

iii) A level-4 logical qubit uses

$$7^4 = 2401$$

physical data qubits. For 10^4 logical qubits, this gives

$$10^4 \cdot 2401 \approx 2.4 \times 10^7$$

physical data qubits.

This is only a back-of-the-envelope estimate. It ignores ancillas for syndrome extraction, verification, routing, idle errors, magic-state distillation, decoding hardware, connectivity constraints, and many architecture-dependent constants. The main lesson is that fault tolerance can suppress logical errors extremely strongly, but only below threshold and with a large physical-qubit overhead.

[3] For part ii), take base-10 logarithms.