The Dangling Information on The Probability Tree

Hammas Hussain Ali (23100047) and Salman Shahid (2310055)

April 2023

1 Abstract

A trip through the forest of classical and quantum information theory is to be undertaken. The notions of information, entropy, and probability will be studied along with their deep and intimate connections. Important quantities in quantum information theory will be studied along with their properties. These quantities include entanglement entropy, average entropy, relative entropy, the Pinsker inequality, mutual information, and correlations. During the trip, the aforementioned concepts shall be applied to a few quantum mechanical systems. The dangling fruit of information on the probability tree will be savored.

2 Introduction

The first report was a mere exhibition of the widely used theorems and concepts in classical and quantum information theory. In this third and final report, we will paint the gallery with proofs for the theorems we discuss, and we will also discuss, in greater depth, the conceptual issues involved. The application to physical systems can be found in the second report.

3 Mathematical formulation

We looked at a few aspects of classical information theory in the first report. We motivated the connection between information and entropy. Following that discussion, one can write an expression for information with a base of e: $H(A) = -\sum_i P(a_i) \ln P(a_i)$ where the entropy is measured in 'nats'. In this report, we will explore classical information theory in greater depth and then move on to quantum information theory. We must realize that information, $H(A)$ is a concave function (this will be motivated further in the section on quantum information). For two events A and B, the information takes the following form:

$$H(A, B) = -\sum_{ij} P(a_i, b_j) \log P(a_i, b_j) \tag{1}$$

This is in terms of a joint probability distribution of measurement outcomes a_i and b_i. Where Individual information for single events A and B can be expressed as

$$\begin{aligned} H(A) &= -\sum_{ij} P(a_i, b_j) \log \sum_k P(a_i, b_k) = -\sum_{ij} P(a_i, b_j) \log P(a_i), \\ H(B) &= -\sum_{ij} P(a_i, b_j) \log \sum_l P(a_l, b_j) = -\sum_{ij} P(a_i, b_j) \log P(b_j). \end{aligned} \tag{2}$$

This is done using

$$\begin{aligned} P(a_i) &= \sum_j P(a_i, b_j), \\ P(b_j) &= \sum_i P(a_i, b_j). \end{aligned} \tag{3}$$

We would now like to find a relation or constraint that connects $H(A), H(B)$, and $H(A, B)$. We shall do this by simplifying the expression for $H(A)+H(B)-H(A, B)$ in the following sense,

$$H(A)+H(B)-H(A, B)=\sum_{ij} P\left(a_i, b_j\right) \log \left(\frac{P\left(a_i, b_j\right)}{P\left(a_i\right) P\left(b_j\right)}\right). \tag{4}$$

Before we move ahead, let us define the notion of relative entropy. Consider two different probability distributions, $P(a_i)$ and $Q(a_i)$, for an event A to have an outcome a_i. Now, we define relative entropy for the two distinct probability distributions such that,

$$H(P \| Q)=\sum_i P\left(a_i\right)\left[\log P\left(a_i\right)-\log Q\left(a_i\right)\right] \tag{5}$$

Where on the left-hand side the double vertical lines just signify that the entropy is relative. In a loose sense, relative entropy allows us to, for example, tell apart a fair coin from a biased one. Both have the same set of outcomes but attach different probabilities to each outcome. Also notice that the way we have defined this quantity, it is not symmetric. That is, $H(P\|Q) \neq H(Q\|P)$. If one looks back at (4), we can write it as relative entropy of the distributions, $\{P\left(a_i, b_j\right)\}$ and $\{P\left(a_i\right) P\left(b_j\right)\}$:

$$\begin{aligned} H(A)+H(B)-H(A, B) & =\sum_{ij} P\left(a_i, b_j\right) \log \left(\frac{P\left(a_i, b_j\right)}{P\left(a_i\right) P\left(b_j\right)}\right) \\ & =H\left(\left\{P\left(a_i, b_j\right)\right\} \|\left\{P\left(a_i\right) P\left(b_j\right)\right\}\right)\end{aligned} \tag{6}$$

We have expressed $H(A), H(B)$, and $H(A, B)$ in terms of relative entropy. Relative entropy is always non-negative. This is a result we proved in class, so we shall omit the proof. Consequently, we have that

$$H(A)+H(B)-H(A, B) \geq 0 \tag{7}$$

This must mean that

$$H(A)+H(B) \geq H(A, B) \tag{8}$$

We might as well define the expression above using a new quantity called mutual information:

$$H(A: B)=H(A)+H(B)-H(A, B) \tag{9}$$

This, as we just saw, must always be non-negative. Mutual information quantifies the "correlations" between events A and B. If A and B are independent, the mutual information just becomes zero. Otherwise, mutual information will be non-zero.

We have built a scaffolding using classical information theory. The time is ripe to erect an edifice of quantum information theory. We will define some interesting concepts and prove some wonderful consequences. It behooves us to begin with Von-Neumann or entanglement entropy (of the subsystem A when the whole system is in the state $|\psi\rangle$):

$$S_A(|\psi\rangle)=-\operatorname{Tr}_A\left(\rho_A \log \left(\rho_A\right)\right) \tag{10}$$

Why is this a viable measure of entanglement? Recall that the degree of entanglement quantifies how much one can "learn" about one system by doing a measurement on another system. It is pertinent to clarify at the onset what we mean by "learning": what new information do we have about this system that we did not have before? Take the bipartite spin state $|01\rangle$ where system A is represented in the first slot, and B in the other. The measurement results obtained from doing a measurement on A tell us nothing that we did not already know about B (i.e. that system B is in the state $|1\rangle$). So this state is not entangled. The same fate befalls all factorized states. Now imagine the state, $\frac{1}{\sqrt{2}}(|01\rangle+|10\rangle)$. This is clearly entangled since doing a measurement on A decides the state for B (either $|0\rangle$ or $|1\rangle$). However, using the logic we have employed until now, we soon run into a problem. Consider the state, $\sqrt{0.9}\,|01\rangle+\sqrt{0.1}\,|10\rangle$ Well, doing a measurement on A still decides something about B but we were already quite sure about the state for A and B so the new measurement does not add much to this. Finally, we expect A to be entangled with B as much as B is entangled with A. We now show that (10) lives by said behavior.

Let's first prove that

$$S_A(|\psi\rangle) = S_B(|\psi\rangle) \tag{11}$$

Consider a bipartite system with subsystems A, B having dimensions dA, dB and orthonormal basis sets, $\{|a_i\rangle\}, \{|b_i\rangle\}$. We begin by writing out a generic state, $|\psi\rangle$ for this system

$$|\psi\rangle = \sum_{i=1}^{dA}\sum_{j=1}^{dB} c_{ij} |a_i\rangle |b_j\rangle . \tag{12}$$

The full density matrix is then

$$|\psi\rangle\langle\psi| = \sum_{ii'}^{dA}\sum_{jj'}^{dB} c^*_{i'j'}c_{ij} |a_i\rangle\langle a'_i| \otimes |b_j\rangle\langle b'_j| . \tag{13}$$

The reduced density matrix for A then becomes

$$\rho_A = \sum_{k}^{dB}\sum_{ii'}^{dA} c^*_{i'k}c_{ik} |a_i\rangle\langle a'_i| \tag{14}$$

Similarly,

$$\rho_B = \sum_{l}^{dA}\sum_{jj'}^{dB} c^*_{lj'}c_{lj} |b_j\rangle\langle b'_j| \tag{15}$$

We now stare at these two expressions above. If we consider c_{ik} as the entries of a dAxdB matrix, C, (14) can be written as

$$\sum_{k}^{dB}\sum_{ii'}^{dA} (C^\dagger)_{ki'}C_{ik} |a_i\rangle\langle a'_i| \tag{16}$$

which lets us (justifiably) claim

$$\rho_A = CC^\dagger \tag{17}$$

Before we proceed, a note of caution: the "=" here is not to be read into; we mean this relation holds in this basis used only. In the same way, ρ_b can be expressed as

$$\sum_{l}^{dA}\sum_{jj'}^{dB} (C^T)_{jl}(C^*)_{lj'} |b_j\rangle\langle b'_j| \tag{18}$$

which is

$$C^TC^* = (C^\dagger C)^*. \tag{19}$$

Since ρ_b is hermitian, its eigenvalues are all real implying that, for the eigenvalue spectrum, we may look at $C^\dagger C$ instead. Now, we will show that both ρ_A and ρ_B have the same non-zero eigenvalues. This proves what we want since, for a n dimensional operator ρ, (10) is a function of its eigenvalues λ_i

$$-\sum_{i=1}^{n} \lambda_i log(\lambda_i) \tag{20}$$

Consider a (dA-dimensional) eigenvector, x, of $CC^\dagger$ with a non-zero eigenvalue, k

$$CC^\dagger x = kx \tag{21}$$

Consider now the dB-dimensional vector $y = C^\dagger x$. This is not the zero vector since neither x nor k are zero. Consider the dance of operators

$$C^\dagger Cy = C^\dagger CC^\dagger x = C^\dagger(CC^\dagger x) = kC^\dagger x = ky \tag{22}$$

We have shown that k is an eigenvalue of $C^\dagger C$. Going in the reverse direction goes the same way. This all but concludes our first proof. Astute readers may have noticed that we brushed aside a crucial detail: we have not said anything about the multiplicity of any of the eigenvalues. That the multiplicities for the non-zero eigenvalues are also the same here is a well-known result in linear algebra [5] and is tedious to repeat here.

We next show that the upper bound is achieved for the types of states mentioned at the onset. To do so, we first find what the upper bound is and then show that said states possess the same. The upper bound of (20) is found under the constraint that $\sum_{i=1}^{n} \lambda_i - 1$. So the relevant function to stare at is

$$h = -\sum_{i=1}^{n} \lambda_i log(\lambda_i) - k(\sum_{i=1}^{n} \lambda_i - 1) \tag{23}$$

where k is our multiplier. Following the usual routine

$$\frac{\partial h}{\partial \lambda_j} = 0 = log\lambda_j - k - 1 \tag{24}$$

which implies

$$\lambda_j = e^{-k-1}. \tag{25}$$

Note how the result is independent of λ_j. All this is to say that our expression achieves its zenith with $\lambda_j = 1/n$ for all j. So ρ_A is of the form

$$\sum_{i=1}^{n} \frac{1}{n} |a_i\rangle \langle a_i| . \tag{26}$$

This corresponds to a maximally mixed state for this subsystem- it behaves exactly like a classic ensemble where there are n distinct outcomes possible with equal likelihood (this relationship between entanglement in the larger state and the subsystem behaving as a classic ensemble is crucial for the study of open quantum systems. This is beyond the scope of our express objectives). What is the maximum entanglement entropy then?

$$-\sum_{i=1}^{n} \frac{1}{n} log\left(\frac{1}{n}\right) = -log\left(\frac{1}{n}\right) = log(n) \tag{27}$$

Recall that n is the dimension of our system. Let's go back to our bipartite system and see what this implies therein. Since $S_A = S_B$ for all states, we must have that

$$S_A, S_B \leq min\,(dA, dB)\,. \tag{28}$$

We now finally consider the (allegedly) maximally entangled state mentioned earlier.

$$a\,|01\rangle + b\,|10\rangle \tag{29}$$

with A in the first slot. We then have

$$\rho_A = a^* a\,|0\rangle \langle 0| + b^* b\,|1\rangle \langle 1| \tag{30}$$

. We can read off the eigenvalues and get the entanglement entropy

$$aa^* log(aa^*) + bb^* log(bb^*). \tag{31}$$

. For $a = b = \frac{1}{\sqrt{2}}$, we get $log(2)$. For $a = \sqrt{0.9}, b = \sqrt{0.1}$, we get $0.47 ln(2)$ as promised. Finally, if the state of the whole system is factorizable, the reduced density matrix is also of a pure state, and a pure state's Von Neumann entropy is zero (to see this, consider $\rho = \rho^2$ for a pure state. This means its eigenvalues are either zero or one). We have finally vindicated this measure of entanglement.

One of our central claims in the previous report was that the entanglement entropy of A with the whole system in the state $|\psi\rangle$, $S_A(|\psi\rangle)$, is a lower bound to the entropy of measurement in A, $S(|\psi\rangle, O_A)$ for all observables, O_A. We now show beyond doubt that this is the case. Recall that

$$S(|\psi\rangle, O_A) = -\sum_{i=1}^{dA} p_i log(p_i) \tag{32}$$

where p_i is the probability of observing the eigenvalue o_i of O_A. This probability is given by

$$Tr(\rho_A |o_i\rangle \langle o_i|) = \langle o_i| \rho_A |o_i\rangle \tag{33}$$

Expanding ρ_A out in its eigenbasis (states $\{|k\rangle\}$ with eigenvalues $\{\lambda_k\}$) gives

$$p_i = \sum_{k=1}^{dA} \lambda_k |\langle o_i|k\rangle|^2 \tag{34}$$

which is just the expectation value of the state $|o_i\rangle$ when the “measurement” ρ_A is carried out on the state $|o_i\rangle$. We now write this out in symbols

$$p_i = \langle\lambda\rangle_i \tag{35}$$

The measurement entropy can then be written as

$$-\sum_{i=1}^{dA} \langle\lambda\rangle_i log\left(\langle\lambda\rangle_i\right) \tag{36}$$

We now make a grand claim

$$-\sum_{i=1}^{dA} \langle\lambda\rangle_i log\left(\langle\lambda\rangle_i\right) \geq -\sum_{i=1}^{dA} \langle\lambda log(\lambda)\rangle_i. \tag{37}$$

We will discuss where this comes from after the larger claim is proven. Now, we write out the right-hand side above using the definition of expectation values

$$-\sum_{i=1}^{dA}\sum_{k=1}^{dA} \lambda_k log(\lambda_k) |\langle o_i|\lambda_k\rangle|^2 \tag{38}$$

The sum can now be written as

$$-\sum_{k=1} \lambda_k log(\lambda_k) \sum_{i=1} |\langle o_i|\lambda_k\rangle|^2 = S_A \tag{39}$$

The last equality follows since the sum over i is one (it is a sum of probabilities of all possible outcomes).

We still have unresolved business: (37). Here we introduce the notion of convex functions. A function is convex in an interval (a, b) if its value in the said interval is always less than or equal to the value taken up by the line connecting $f(a)$ and $f(b)$. A way to mathematically quantify what we have said is if f is convex, then

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \tag{40}$$

for $\alpha \in [0, 1]$ and $x_1, x_2 \in (a, b)$. To see that this is indeed what we said above, note that if we imagine α running from 1 to 0, the inputs to the function on the left go from x_1 to x_2. Moreover, the equation on the right is the parametric equation of the line connecting $(x_1, f(x_1))$ and $(x_2, f(x_2))$ evaluated at the input to the function on the left side. Now stare at this: think of α as the probability of outcome x_1 and $1-\alpha$ for x_2. From this, we motivate Jensen's inequality for a convex function f:

$$f\left(\sum_{i}^{n} p_i x_i\right) \leq \sum_{i}^{n} p_i f(x_i). \tag{41}$$

Note that the input to the function on the left is the expectation value of x and on the right, we have the expectation value of $f(x)$. In our proof above, we had the function $\lambda log(\lambda)$. Is it convex? If a function has a second derivative, another equivalent characterization of convexity is that this second derivative is non-negative; our function meets this condition. Note, however, we used the inequality swapped by multiplying a negative sign. Another way to think about this is that while $\lambda log(\lambda)$ is convex, $-\lambda log(\lambda)$ is concave. While this is mathematically obvious from the aforementioned, this fits in nicely physically. Our classical definition of information is a sum of such functions, $-\sum_i P(a_i) \ln P(a_i)$. So we expect it to be concave too. For simplicity, we can think of a system with two possible outcomes with probabilities p and $1-p$. We expect the function to be peaked when $p = 0.5$ since we are most uncertain (and hence can learn the most) then. Crucially, any change in p from this value will lead to a decline in the information content. The function will essentially look like an upside-down U, which is concave - as we had mathematically shown it must be since it is a sum of concave functions.

We now move to a definition of the quantum relative entropy in a fashion similar to its classical counterpart:

$$S(\hat{\sigma}\|\hat{\rho}) = \mathrm{Tr}\left(\hat{\sigma}(\log\hat{\sigma} - \log\hat{\rho})\right) \tag{42}$$

Where $\hat{\rho}$ and $\hat{\sigma}$ are two density operators. The expression above gives us a measure of the "distinguishability" of two density operators. Notice that if in the expression above the density operators are same, the relative entropy vanishes. We will now attempt to prove the lower bound of the quantum relative entropy. We begin by writing the aforementioned density operators in their diagonal form. We can always do this because density operators, by definition, are hermitian and must be diagonalizable:

$$\begin{aligned} \hat{\rho} &= \sum_m \rho_m \left|\rho_m\right\rangle \left\langle\rho_m\right| \\ \hat{\sigma} &= \sum_n \sigma_n \left|\sigma_n\right\rangle \left\langle\sigma_n\right| \end{aligned} \tag{43}$$

We have diagonalized the density operators using the orthonormal eigenbasis of σ and ρ. We will write down relative entropy in terms of diagonalized density operators:

$$S(\hat{\sigma}\|\hat{\rho}) = \sum_n \sigma_n \log\sigma_n - \sum_{nm} \sigma_n \left|\left\langle\sigma_n \mid \rho_m\right\rangle\right|^2 \log\rho_m \tag{44}$$

This follows because

$$Tr[\sum_n \sigma_n \left|\sigma_n\right\rangle \left\langle\sigma_n\right| log(\sum_n \sigma_n \left|\sigma_n\right\rangle \left\langle\sigma_n\right|)] = \sum_n \sigma_n \log\sigma_n \tag{45}$$

and

$$Tr\left[\sum_n \sigma_n \left|\sigma_n\right\rangle \left\langle\sigma_n\right| log(\sum_m \rho_m \left|\rho_m\right\rangle \left\langle\rho_m\right|)\right] = \sum_{nm} \sigma_n \left|\left\langle\sigma_n \mid \rho_m\right\rangle\right|^2 \log\rho_m \tag{46}$$

We can show both results from above using arguments similar to those covered in class. Now, remember that we have a completeness relation over the eigenstates of ρ:

$$\sum_m \left|\rho_m\right\rangle \left\langle\rho_m\right| = 1 = \sum_m \left|\left\langle\sigma_n \mid \rho_m\right\rangle\right|^2 = 1 \tag{47}$$

We can use this to get

$$S(\hat{\sigma}\|\hat{\rho}) = \sum_{nm} \sigma_n \left|\left\langle\sigma_n \mid \rho_m\right\rangle\right|^2 (\log\sigma_n - \log\rho_m) \tag{48}$$

The expression above can be written as

$$\begin{aligned} &\sum_{nm} \sigma_n \left|\left\langle\sigma_n \mid \rho_m\right\rangle\right|^2 \Big(\log\sigma_n + \log\left|\left\langle\sigma_n \mid \rho_m\right\rangle\right|^2 \\ &\qquad - \log\rho_m - \log\left|\left\langle\sigma_n \mid \rho_m\right\rangle\right|^2\Big) \end{aligned} \tag{49}$$

where we add and subtract the $\log |\langle \sigma_n \mid \rho_m \rangle|^2$ term. We can combine the logarithm terms now to get,

$$\sum \sigma_n |\langle \sigma_n \mid \rho_m \rangle|^2 \left[\log \left(\sigma_n |\langle \sigma_n \mid \rho_m \rangle|^2 \right) - \log \left(\rho_m |\langle \sigma_n \mid \rho_m \rangle|^2 \right) \right] \tag{50}$$

Now, notice that the $\sigma_n |\langle \sigma_n \mid \rho_m \rangle|^2$ term can not be negative because σ_n has to be positive because of density operators must be semi-definite. The above-mentioned term also adds up to zero if one sums over n and m because of the completeness relation of the ρ and the trace of σ must be one. That must mean that $\sigma_n |\langle \sigma_n \mid \rho_m \rangle|^2$ and $\rho_m |\langle \sigma_n \mid \rho_m \rangle|^2$ are probabilities $P(n,m)$ and $Q(n,m)$ respectively. Consequently, we have that

$$S(\hat{\sigma} \| \hat{\rho}) = \sum_{nm} P(n,m)[\log P(n,m) - \log Q(n,m)] \tag{51}$$

This is surprisingly just like the expression of classical relative entropy from (5). This as mentioned above must be greater than or equal to zero. Hence, (51) must be greater than or equal to zero. Hence, the quantum relative entropy must be greater than or equal to zero!

Let us discuss how relative entropy is an upper bound for an important quantity: the normalized difference of expectation values of density operators. Let us present to you the result in its full glory and then explain its components and furnish you with a proof.

$$\frac{1}{2||O||^2}(tr(O\rho) - tr(O\sigma))^2 \leq S(\rho \| \sigma) \tag{52}$$

Where $\hat{O}$ are bounded operators, σ and ρ are density operators. The bound above must be true for all bounded operators. $\|O\|$ is the norm of the operator. To understand the various components of the inequality above, we start with the notion of an operator and bounded operators. Consider two normed vector spaces X and Y. In our case, X and Y are Hilbert spaces as well. Suppose that O is a map between the two vector spaces:

$$O : X \to Y \tag{53}$$

This map, O, must be well-behaved (continuous and linear). We define a quantity called the operator norm in the following way

$$\|O\| = \sup_{|\psi\rangle} \sqrt{\frac{\langle \psi | O^\dagger O | \psi \rangle}{\langle \psi \mid \psi \rangle}} \tag{54}$$

Where sup is the supremum over the vectors, $|\psi\rangle$, that belong to the vector space X. If $\|O\| < \infty$, O is called a bounded operator. Notice that in the definition of the operator norm above, we are essentially, considering the ratio of the "size" of a vector in vector space Y to the size of a vector in vector space X.

To prove the inequality from (52), we need to prove a few other lemmas. Let us state and prove those and come back to the original inequality. The first lemma is

$$S(\rho \| \sigma) \geq \frac{1}{2ln(2)} (\|\rho - \sigma\|_1)^2 \tag{55}$$

This is called the Pinsker inequality and serves as an upper bound on the trace norm. The trace norm is defined by

$$\|X\|_1 = \mathrm{Tr}(\sqrt{X^\dagger X}) \tag{56}$$

Where X is an operator. For the Pinsker inequality, the trace norm of interest is

$$\|\rho - \sigma\|_1 = \mathrm{Tr}\left(\sqrt{(\rho - \sigma)^\dagger (\rho - \sigma)} \right) \tag{57}$$

We will attempt to prove the Pinsker inequality for qubit states ρ and σ that are diagonal in the same basis. That is, they commute because they share an eigenbasis. We have that,

$$\begin{aligned} \rho &\equiv p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| \\ \sigma &\equiv q|0\rangle\langle 0| + (1-q)|1\rangle\langle 1| \end{aligned} \tag{58}$$

Without loss of generality, we assume that $p \geq q$. Notice that

$$\begin{aligned}S(\rho \mid \sigma) &= plogp + (1-p)log(1-p) - plogq - (1-p)log(1-q) \\ &= plog\left(\frac{p}{q}\right) + (1-p)log[(1-p)(1-q)]\end{aligned} \tag{59}$$

Also, notice that

$$(\|\rho-\sigma\|_1) = Tr\left(\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}\right)$$
$$= Tr\left(\sqrt{p^2|0\rangle\langle 0| + (1-p)^2|1\rangle\langle 1| + q^2|0\rangle\langle 0| + (1-q)^2|1\rangle\langle 1| - 2pq|0\rangle\langle 0| - 2(1-p)(1-q)|1\rangle\langle 1|}\right) \tag{60}$$

After considering the positive square root in the expression above, we have

$$Tr[(p-q)|0\rangle\langle 0| + (p-q)|1\rangle\langle 1|] \tag{61}$$

After the trace we get,

$$2(p-q) \implies (\|\rho-\sigma\|_1)^2 = 4(p-q)^2 \tag{62}$$

Let us construct a function $g(p,q)$ in the following manner

$$g(p,q) \equiv p\log\left(\frac{p}{q}\right) + (1-p)\log\left(\frac{1-p}{1-q}\right) - \frac{4}{2\ln 2}(p-q)^2 \tag{63}$$

Proving (55) for commuting ρ, σ is then tantamount to proving that $g(p,q)$ is non-negative for all relevant values of p, q. To do so, we show that $\frac{\partial g(p,q)}{\partial q}$ is less than or equal to zero.

$$\begin{aligned}\frac{\partial g(p,q)}{\partial q} &= -\frac{p}{q\ln 2} + \frac{1-p}{(1-q)\ln 2} - \frac{4}{\ln 2}(q-p) \\ &= -\frac{p(1-q)}{q(1-q)\ln 2} + \frac{q(1-p)}{q(1-q)\ln 2} - \frac{4}{\ln 2}(q-p) \\ &= \frac{q-p}{q(1-q)\ln 2} - \frac{4}{\ln 2}(q-p) \\ &= \frac{(q-p)\left(4q^2-4q+1\right)}{q(1-q)\ln 2} \\ &= \frac{(q-p)(2q-1)^2}{q(1-q)\ln 2} \\ &\leq 0,\end{aligned} \tag{64}$$

We can carry out the final step because $p \geq q$ and $1 \geq q \geq 0$. Imagine that we fix some $0 \leq p \leq 1$ and vary q (with the restriction that $p \geq q$) as before. When $p = q$, we have that $g(p,q) = 0$: the function decreases or stays the same as q is varied from 0 to p and reaches its minimum when $p = q$ (for the interval where $p \geq q$). We are confident that this is the first time the function hits 0 since had it passed below the x-axis before this point, there would be no way to "pull back up" since it must either remain the same or keep decreasing. Thus, $g(p,q)$ is indeed non-negative in the region of interest.

Our next task is to generalize the above for ρ, σ that do not commute. Consider the hermitian operator $\rho - \sigma$. We can then construct Π, the projector into this operator's positive eigenspace (by which we just mean the subspace of the Hilbert space spanned by the eigenbasis of all non-negative eigenvalues). Let us now define an operation, $\mathcal{M}$, on ρ, σ as follows

$$\begin{aligned}\mathcal{M}(\rho) &= \mathrm{Tr}\{\Pi\rho\}|0\rangle\langle 0| + \mathrm{Tr}\{(I-\Pi)\rho\}|1\rangle\langle 1|, \\ \mathcal{M}(\sigma) &= \mathrm{Tr}\{\Pi\sigma\}|0\rangle\langle 0| + \mathrm{Tr}\{(I-\Pi)\sigma\}|1\rangle\langle 1|.\end{aligned} \tag{65}$$

To make a connection to what we proved before, let $p = \mathrm{Tr}\{\Pi\rho\}$ and $q = \mathrm{Tr}\{\Pi\sigma\}$. We are confident that both are positive and less than or equal to one since they represent probabilities (of landing in the positive eigenspace given the state ρ or σ). Now either $p \geq q$ or $p < q$. Either way, the connection to what was done before stands (by swapping p and q as needed). So we are then able to say

$$S(\mathcal{M}\rho\|\mathcal{M}\sigma) \geq \frac{4}{2\ln 2}(p-q)^2 = \frac{4}{2\ln 2}(\mathrm{Tr}\{\Pi\rho\} - \mathrm{Tr}\{\Pi\sigma\})^2 = \frac{1}{2\ln 2}(2\,\mathrm{Tr}\{\Pi(\rho-\sigma)\})^2 \tag{66}$$

Now we make two claims that are justified after the express proof is complete. First, we claim that this last expression above is equal to $\frac{1}{2\ln 2}||\rho-\sigma||^2$ yielding

$$S(\mathcal{M}\rho\|\mathcal{M}\sigma) \geq \frac{1}{2\ln 2}||\rho-\sigma||^2. \tag{67}$$

Second, we claim that $\mathcal{M}$ is a map such that

$$S(\rho\|\sigma) \geq S(\mathcal{M}\rho\|\mathcal{M}\sigma). \tag{68}$$

In consequence of the foregoing, we finally have

$$S(\rho\|\sigma) \geq \frac{1}{2\ln 2}||\rho-\sigma||^2. \tag{69}$$

We now attempt to substantiate the two claims. For the first claim, note that $\rho-\sigma$ is hermitian and so is diagonalizable. We can split the diagonal matrix into two parts: one with the non-negative eigenvalues and one with the absolute value of the negative eigenvalues in their correct positions on the diagonal

$$\rho - \sigma = D^+ - D^- \tag{70}$$

where D^+, D^- are matrices written in the eigenbasis of $\rho-\sigma$. We can also construct the projector onto the non-negative and negative eigenspaces, Π^+ and $\Pi^- = \mathbb{1} - \Pi^+$ respectively. Now,

$$||\rho-\sigma|| = \mathrm{Tr}\{|\rho-\sigma|\}. \tag{71}$$

But

$$|\rho-\sigma| = |D^+ - D^-| = D^+ + D^-. \tag{72}$$

Here the sign change is simply because the non-negative and negative eigenvalues are on different diagonal entries. This yields

$$||\rho-\sigma|| = \mathrm{Tr}\{D^+ + D^-\} = \mathrm{Tr}\{D^+\} + \mathrm{Tr}\{D^-\} \tag{73}$$

Consider now

$$\mathrm{Tr}\{D^+\} - \mathrm{Tr}\{D^-\} = \mathrm{Tr}\{D^+ - D^-\} = \mathrm{Tr}\{\rho-\sigma\} = \mathrm{Tr}\{\rho\} - \mathrm{Tr}\{\sigma\} = 1 \tag{74}$$

where the last equality follows since the trace of all density matrices is one. The last two results then lead to

$$||\rho-\sigma|| = 2\,\mathrm{Tr}\{D^+\}. \tag{75}$$

Finally, putting this altogether gives

$$\begin{aligned} 2\,\mathrm{Tr}\left\{\Pi^+(\rho-\sigma)\right\} &= 2\,\mathrm{Tr}\left\{\Pi^+\left(D^+ - D^-\right)\right\} \\ &= 2\,\mathrm{Tr}\left\{\Pi^+ D^+\right\} \\ &= 2\,\mathrm{Tr}\left\{D^+\right\} \\ &= \|\rho-\sigma\|_1. \end{aligned} \tag{76}$$

One way to show that we can ignore $\Pi^+ D^-$ is to write out D^- as $\Pi^- D^-$ (recall that Π^- projects into the negative eigenspace and D^- has ones in the diagonal for the basis of that eigenspace) and then note that $\Pi^+\Pi^- D^- = 0$ since $\Pi^+\Pi - = 0$ since they are projectors into spaces orthogonal to each other. By an argument similar to the former, we also conclude that $\Pi^+D^+ = D^+$. Using these two arguments, the second and third equalities in the sequence above follow.

The second claim rests on the well-known inequality: the Monotonicity of quantum relative entropy. It claims that

$$S(\rho\|\sigma) \geq S(\mathcal{M}\rho\|\mathcal{M}\sigma) \tag{77}$$

where $\mathcal{M}$ is a completely positive trace-preserving operation. We will not prove this claim but will establish that the particular operation we employed in our proof fits the bill.

Let's check trace preservation first. We can read off $\mathrm{Tr}\{\mathcal{M}\rho\}$ from (65)

$$\mathrm{Tr}\{\Pi\rho\} + \mathrm{Tr}\{(\mathbb{1} - \Pi)\rho\} = \mathrm{Tr}\{\rho\} \tag{78}$$

So $\mathcal{M}$ conserves the trace.

Positivity means that if $\mathcal{M}$ acts on a positive operator, the output is also a positive operator. We consider the action of $\mathcal{M}$ on density operators (which are positive). We have already argued that $\mathrm{Tr}\{\Pi\rho\}, \mathrm{Tr}\{(\mathbb{1} - \Pi)\rho\}$ are probabilities and so are semi-positive. Since the eigenvalues of $\mathcal{M}\rho$ are greater than or equal to zero, it is a positive operator. But does this hold for all positive operators? Since we are in a complex vector space, any positive operator is also Hermitian. Consider writing any generic positive operator down in its diagonal form. If we multiply this with the reciprocal of its trace, we achieve a viable density matrix. As such, we can write any positive operator in terms of a density operator, and our proof above kicks in.

Complete positivity means that $\mathcal{M}$ preserves positivity even with auxiliary spaces. To understand this, create a tensor product operator $\mathbb{1}_{k\times k} \otimes \mathcal{M}$. If our map is completely positive then the action of this tensor product operator on $A \otimes \rho$, where A is an arbitrary k-dimensional operator and ρ is a positive operator, produces another positive operator. Unfortunately, at the time of writing, we are unable to show that $\mathcal{M}$ fits this description.

We now attempt to prove (52). To do so, we need an inequality that lets us introduce the arbitrary operator, O found in the expression. We state such an inequality.

$$\|X\|_1 \geq \mathrm{Tr}\left(\frac{XO}{\|O\|}\right) \tag{79}$$

Replacing X with $\rho - \sigma$ gives

$$\|\rho - \sigma\|_1 \geq \mathrm{Tr}\left(\frac{(\rho - \sigma)O}{\|O\|}\right) \tag{80}$$

We can now square both sides. If the right-hand side is negative making squaring illegal, consider $(\sigma - \rho)$ instead. Note that the left-hand side is the same for both. So we then have

$$(\|\rho - \sigma\|_1)^2 \geq \left(\mathrm{Tr}\left(\frac{(\rho - \sigma)O}{\|O\|}\right)\right)^2 \tag{81}$$

This read together with the Pinsker inequality yields the desired result.

4 Implementation methodology

We have acquired the tools from theory up till now. It is now time to use some of the concepts and tools that have been discussed and apply them to a physical system. The system we are interested in is one with two interacting spin-half systems, A, B, in an external magnetic field $\vec{B}$. We will start with an un-entangled state and produce an entangled one through time evolution. In our example, the entanglement entropy varies with time. The foregoing raises a question: what is the average entropy of the state as it evolves? Our states evolves into different parts of the Hilbert Space and the average calculated is over this collection of states. This is the average entropy over time. However, there is another average that merits discussion, especially vis-a-vis the former. We consider an ensemble of random states spread all over the Hilbert space and calculate their average. This is the average entropy over an ensemble of states. We shall discuss both of these averages.

4.1 Time evolution of entanglement entropy

Let us orient the magnetic field in the z-axis only. The expanded Hamiltonian is

$$H = \vec{B}.\vec{J_A} + \vec{B}.\vec{J_B} + g\vec{J_B}.\vec{J_A} \propto J_A^z + J_B^z + g'\vec{J_B}.\vec{J_A} \tag{82}$$

where $\vec{J_A}$ and $\vec{J_B}$ are spin operators for systems A and B, and g' plays the role of a time-dependent coupling constant. The constant,g', is non-zero initially and becomes zero at some final time. Consequently, we have the following the unitary time evolution operation

$$U = e^{-iHt} \tag{83}$$

Where H is a time-independent Hamiltonian. Let us consider the initial state of the form $|\psi_o\rangle = |\uparrow\downarrow\rangle$. This state must evolve in time because of the action of the unitary time evolution operator. The first Mathematica notebook submitted in the second report will implement this system and plot the entanglement entropy, $-Tr(\rho log(\rho))$. The result of this can be found in figure (1) from section 5. The code for this is also included in Notebook 1 with a detailed walk-through. The code starts with setting up the states, Pauli matrices, etc. Then we create an entangled ground state of the Hamiltonian from (82). This turns out to be the singlet state. The singlet state is shown to have maximum ($log(2)$) entanglement entropy. We then set up the system for (1) and employ it to evolve a factorized state (which, of course, has zero entanglement entropy). This setup is used to plot the entanglement entropy of said evolving state in three stages: before the interaction turns on, while it is on, and finally after the interaction is turned off. It is pertinent to point out here that for the third stage - the post-interaction period- we plot the final entanglement entropy of stage 2 as the constant value therein. Of course, this need not be the case if the state at the end of stage 2 is not an eigenstate of the non-interacting Hamiltonian. However, our evolving state is at all times an eigenstate of it.

4.2 Entanglement entropy distribution for an ensemble

Now, we will make a random ensemble of states by selecting from a normal distribution. Then we will calculate the entanglement entropy of the random states in the ensemble and plot the frequency of a particular value of entanglement entropy. The result of this can be found in figure (2) from section 3. The code for this can be found in the second Mathematica Notebook along with a detailed walk-through. We start off with creating a bipartite system with N number of spin-1/2 systems distributed into subsystems, A and B. We create random states by selecting coefficients of a state from a normal distribution of mean zero and standard deviation of one. We create multiple random states to create an ensemble. Now, we make a reduced density matrix for each of these random states. For each of these density matrices, we find the corresponding entanglement entropies. These values are then plotted and compared to the maximum entanglement entropy. We obtain a distribution that resembles a Gaussian.

5 Results

5.1 Time evolution of entanglement entropy

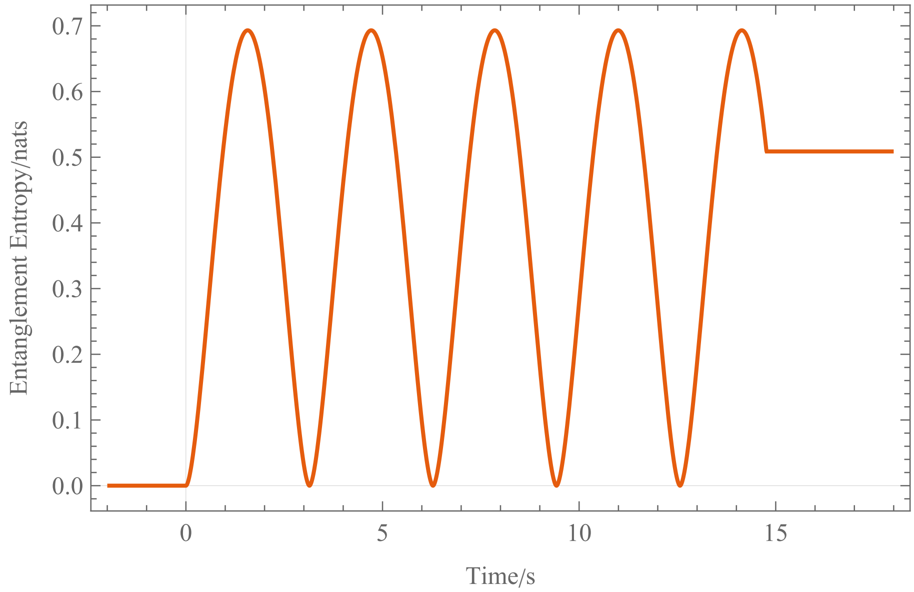

Figure 1: Time evolution of entanglement entropy for a factorized initial state

5.2 Entanglement entropy of an ensemble

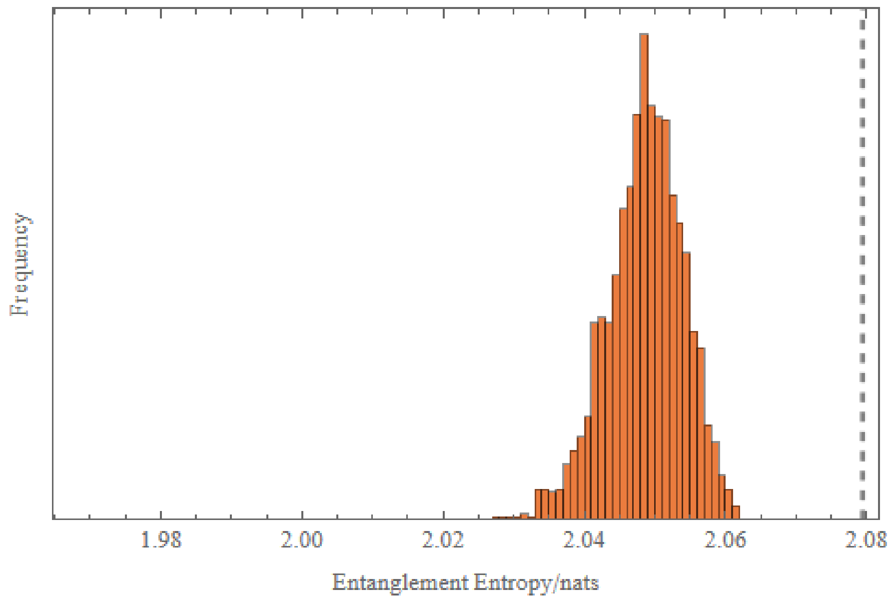

Figure 2: Entanglement entropy distribution for an ensemble

6 Conclusion

It is time now to bid farewell. We started the voyage with classical information theory. We visited various trees to learn about plethora of concepts in classical information. These included the relation between probability and information, mutual information, and relative entropy. We looked at these concepts at an intuitive level and also proved various theorems and bounds for the aforementioned quantities. We then came to a forest of quantum information theory. We looked at quantities analogous to the ones from classical information theory. Various results in quantum information theory were proved and an interpretation was provided for them. This report ended with applying various tools from quantum information to a concrete physical system. A code for computing and plotting various quantities related to the system of interest was written down. The resulting plots were also provided in this report. There is much more to the information theory landscape; the forest is too vast to be mapped out in one report.

References

[1] Quantum Information Theory - Mark M. Wilde

[2] Quantum Information Stephen - M. Barnett

[3] Quantum Information for Quantum Gravity: Selected Topics - Lecture 1 - Eugenio Bianchi

[4] Quantum Information for Quantum Gravity: Selected Topics - Lecture 2 - Eugenio Bianchi

[5] Block Matrices in Linear Algebra - Stephan R. Garcia