CS 766/QIC 820 Theory of Quantum Information (Fall 2011)

Lecture 10: Continuity of von Neumann entropy; quantum relative entropy

In the previous lecture we defined the Shannon and von Neumann entropy functions, and established the fundamental connection between these functions and the notion of compression. In this lecture and the next we will look more closely at the von Neumann entropy in order to establish some basic properties of this function, as well as an important related function called the *quantum relative entropy*.

10.1 Continuity of von Neumann entropy

The first property we will establish about the von Neumann entropy is that it is continuous everywhere on its domain.

First, let us define a real valued function $\eta : [0, \infty) \to \mathbb{R}$ as follows:

$$\eta(\lambda) = \begin{cases} -\lambda \ln(\lambda) & \lambda > 0 \\ 0 & \lambda = 0. \end{cases}$$

This function is continuous everywhere on its domain, and derivatives of all orders exist for all positive real numbers. In particular we have $\eta'(\lambda) = -(1 + \ln(\lambda))$ and $\eta''(\lambda) = -1/\lambda$. A plot of the function η is shown in Figure 10.1, and its first derivative η' is plotted in Figure 10.2.

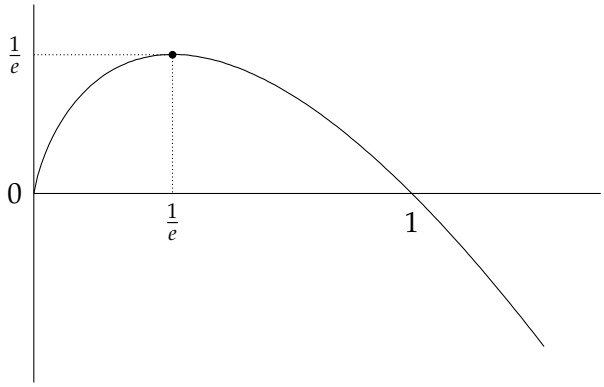

Figure 10.1: A plot of the function $\eta(\lambda) = -\lambda \ln(\lambda)$.

The fact that η is continuous on $[0, \infty)$ implies that for every finite, nonempty set Σ the Shannon entropy is continuous at every point on $[0, \infty)^{\Sigma}$, as

$$H(p) = \frac{1}{\ln(2)} \sum_{a \in \Sigma} \eta(p(a)).$$

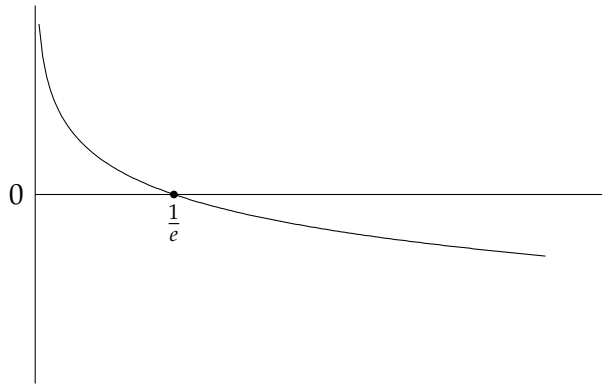

Figure 10.2: A plot of the function $\eta'(\lambda) = -(1 + \ln(\lambda))$.

We are usually only interested in $H(p)$ for probability vectors p, but of course the function is defined on vectors having nonnegative real entries.

Now, to prove that the von Neumann entropy is continuous, we will first prove the following theorem, which establishes one specific sense in which the eigenvalues of a Hermitian operator vary continuously as a function of an operator. We don't really need the precise bound that this theorem establishes—all we really need is that eigenvalues vary continuously as an operator varies, which is somewhat easier to prove and does not require Hermiticity—but we'll take the opportunity to state the theorem because it is interesting in its own right.

Theorem 10.1. *Let $\mathcal{X}$ be a complex Euclidean space and let $A, B \in \mathrm{Herm}(\mathcal{X})$ be Hermitian operators. It holds that*

$$\|\lambda(A) - \lambda(B)\|_1 \leq \|A - B\|_1 .$$

To prove this theorem, we need another fact about eigenvalues of operators, but this one we will take as given. (You can find proofs in several books on matrix analysis.)

Theorem 10.2 (Weyl's monotonicity theorem)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $A, B \in \mathrm{Herm}(\mathcal{X})$ satisfy $A \leq B$. It holds that $\lambda_j(A) \leq \lambda_j(B)$ for $1 \leq j \leq \dim(\mathcal{X})$.*

Proof of Theorem 10.1. Let $n = \dim(\mathcal{X})$. Using the spectral decomposition of $A - B$, it is possible to define two positive semidefinite operators $P, Q \in \mathrm{Pos}(\mathcal{X})$ such that:

1. $PQ = 0$, and
2. $P - Q = A - B$.

(An expression of a given Hermitian operator as $P - Q$ for such a choice of P and Q is sometimes called a *Jordan–Hahn decomposition* of that operator.) Notice that $\|A - B\|_1 = \mathrm{Tr}(P) + \mathrm{Tr}(Q)$.

Now, define one more Hermitian operator

$$X = P + B = Q + A.$$

We have $X \geq A$, and therefore $\lambda_j(X) \geq \lambda_j(A)$ for $1 \leq j \leq n$ by Weyl's monotonicity theorem. Similarly, it holds that $\lambda_j(X) \geq \lambda_j(B)$ for $1 \leq j \leq n = \dim(\mathcal{X})$. By considering the two possible cases $\lambda_j(A) \geq \lambda_j(B)$ and $\lambda_j(A) \leq \lambda_j(B)$, we therefore find that

$$\big|\lambda_j(A) - \lambda_j(B)\big| \leq 2\lambda_j(X) - (\lambda_j(A) + \lambda_j(B))$$

for $1 \leq j \leq n$. Thus,

$$\|\lambda(A) - \lambda(B)\|_1 = \sum_{j=1}^{n} \big|\lambda_j(A) - \lambda_j(B)\big| \geq \mathrm{Tr}(2X - A - B) = \mathrm{Tr}(P + Q) = \|A - B\|_1$$

as required. □

With the above fact in hand, it is immediate from the expression $S(P) = H(\lambda(P))$ that the von Neumann entropy is continuous (as it is a composition of two continuous functions).

Theorem 10.3. *For every complex Euclidean space $\mathcal{X}$, the von Neumann entropy $S(P)$ is continuous at every point $P \in \mathrm{Pos}(\mathcal{X})$.*

Let us next prove Fannes' inequality, which may be viewed as a quantitative statement concerning the continuity of the von Neumann entropy. To begin, we will use some basic calculus to prove a fact about the function η.

Lemma 10.4. *Suppose α and β are real numbers satisfying $0 \leq \alpha \leq \beta \leq 1$ and $\beta - \alpha \leq 1/2$. It holds that*

$$|\eta(\beta) - \eta(\alpha)| \leq \eta(\beta - \alpha).$$

Proof. Consider the function $\eta'(\lambda) = -(1 + \ln(\lambda))$, which is plotted in Figure 10.2. Given that η' is monotonically decreasing on its domain $(0, \infty)$, it holds that the function

$$f(\lambda) = \int_{\lambda}^{\lambda+\gamma} \eta'(t)dt = \eta(\lambda + \gamma) - \eta(\lambda)$$

is monotonically non-increasing for any choice of $\gamma \geq 0$. This means that the maximum value of $|f(\lambda)|$ over the range $\lambda = [0, 1 - \gamma]$ must occur at either $\lambda = 0$ or $\lambda = 1 - \gamma$, and so for λ in this range we have

$$|\eta(\lambda + \gamma) - \eta(\lambda)| \leq \max\{\eta(\gamma), \eta(1 - \gamma)\}.$$

Here we have used the fact that $\eta(1) = 0$ and $\eta(\lambda) \geq 0$ for $\lambda \in [0, 1]$.

To complete the proof it suffices to prove that $\eta(\gamma) \geq \eta(1 - \gamma)$ for $\gamma \in [0, 1/2]$. This claim is certainly supported by the plot in Figure 10.1, but we can easily prove it analytically. Define a function $g(\lambda) = \eta(\lambda) - \eta(1 - \lambda)$. We see that g happens to have zeroes at $\lambda = 0$ and $\lambda = 1/2$, and were there an additional zero λ of g in the range $(0, 1/2)$, then we would have two distinct values $\delta_1, \delta_2 \in (0, 1/2)$ for which $g'(\delta_1) = g'(\delta_2) = 0$ by the mean value theorem. This, however, is in contradiction with the fact that the second derivative $g''(\lambda) = \frac{1}{1-\lambda} - \frac{1}{\lambda}$ of g is strictly negative in the range $(0, 1/2)$. As $g(1/4) > 0$, for instance, we have that $g(\lambda) \geq 0$ for $\lambda \in [0, 1/2]$ as required. □

Theorem 10.5 (Fannes Inequality)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $n = \dim(\mathcal{X})$. For all density operators $\rho, \xi \in \mathrm{D}(\mathcal{X})$ such that $\|\rho - \xi\|_1 \leq 1/e$ it holds that*

$$|S(\rho) - S(\xi)| \leq \log(n) \|\rho - \xi\|_1 + \frac{1}{\ln(2)} \eta(\|\rho - \xi\|_1).$$

Proof. Define

$$\varepsilon_i = |\lambda_i(\rho) - \lambda_i(\xi)|$$

and let $\varepsilon = \varepsilon_1 + \cdots + \varepsilon_n$. Note that $\varepsilon_i \leq \|\rho - \xi\|_1 \leq 1/e < 1/2$ for each i, and therefore

$$|S(\rho) - S(\xi)| = \frac{1}{\ln(2)} \left| \sum_{i=1}^{n} \eta(\lambda_i(\rho)) - \eta(\lambda_i(\xi)) \right| \leq \frac{1}{\ln(2)} \sum_{i=1}^{n} \eta(\varepsilon_i)$$

by Lemma 10.4.

For any positive α and β we have $\beta\eta(\alpha/\beta) = \eta(\alpha) + \alpha\ln(\beta)$, so

$$\frac{1}{\ln(2)} \sum_{i=1}^{n} \eta(\varepsilon_i) = \frac{1}{\ln(2)} \sum_{i=1}^{n} (\varepsilon\eta(\varepsilon_i/\varepsilon) - \varepsilon_i \ln(\varepsilon)) = \frac{\varepsilon}{\ln(2)} \sum_{i=1}^{n} \eta(\varepsilon_i/\varepsilon) + \frac{1}{\ln(2)}\eta(\varepsilon).$$

Because $(\varepsilon_1/\varepsilon, \ldots, \varepsilon_n/\varepsilon)$ is a probability vector this gives

$$|S(\rho) - S(\xi)| \leq \varepsilon \log(n) + \frac{1}{\ln(2)}\eta(\varepsilon).$$

We have that $\varepsilon \leq \|\rho - \xi\|_1$, and that η is monotone increasing on the interval $[0, 1/e]$, so

$$|S(\rho) - S(\xi)| \leq \log(n) \|\rho - \xi\|_1 + \frac{1}{\ln(2)}\eta(\|\rho - \xi\|_1),$$

which completes the proof. □

10.2 Quantum relative entropy

Next we will introduce a new function, which is indispensable as a tool for studying the von Neumann entropy: the *quantum relative entropy*. For two positive definite operators $P, Q \in \mathrm{Pd}(\mathcal{X})$ we define the quantum relative entropy of P with Q as follows:

$$S(P\|Q) = \mathrm{Tr}(P\log(P)) - \mathrm{Tr}(P\log(Q)). \tag{10.1}$$

We usually only care about the quantum relative entropy for density operators, but there is nothing that prevents us from allowing the definition to hold for all positive definite operators.

We may also define the quantum relative entropy for positive semidefinite operators that are not positive definite, provided we are willing to have an extended real-valued function. Specifically, if there exists a vector $u \in \mathcal{X}$ such that $u^*Qu = 0$ and $u^*Pu \neq 0$, or (equivalently) when

$$\ker(Q) \not\subseteq \ker(P),$$

we define $S(P\|Q) = \infty$. Otherwise, there is no difficulty in evaluating the above expression (10.1) by following the usual convention of setting $0\log(0) = 0$. Nevertheless, it will typically not be necessary for us to give up the convenience of restricting our attention to positive definite operators. This is because we already know that the von Neumann entropy function is continuous, and we will mostly use the quantum relative entropy in this course to establish facts about the von Neumann entropy.

The quantum relative entropy $S(P\|Q)$ can be negative for some choices of P and Q, but not when they are density operators (or more generally when $\mathrm{Tr}(P) = \mathrm{Tr}(Q)$). The following theorem establishes that this is so, and in fact that the value of the quantum relative entropy of two density operators is zero if and only if they are equal.

Theorem 10.6. *Let $\rho, \xi \in \mathrm{D}(\mathcal{X})$ be positive definite density operators. It holds that*

$$S(\rho\|\xi) \geq \frac{1}{2\ln(2)} \|\rho - \xi\|_2^2.$$

Proof. Let us first note that for every choice of $\alpha, \beta \in (0,1)$ we have

$$\alpha \ln(\alpha) - \alpha \ln(\beta) = (\alpha - \beta)\eta'(\beta) + \eta(\beta) - \eta(\alpha) + \alpha - \beta.$$

Moreover, by Taylor's Theorem, we have that

$$(\alpha - \beta)\eta'(\beta) + \eta(\beta) - \eta(\alpha) = -\frac{1}{2}\eta''(\gamma)(\alpha - \beta)^2$$

for some choice of γ lying between α and β.

Now, let $n = \dim(\mathcal{X})$ and let

$$\rho = \sum_{i=1}^{n} p_i x_i x_i^* \quad \text{and} \quad \xi = \sum_{i=1}^{n} q_i y_i y_i^*$$

be spectral decompositions of ρ and ξ. The assumption that ρ and ξ are positive definite density operators implies that p_i and q_i are positive for $1 \leq i \leq n$. Applying the facts observed above, we have that

$$\begin{aligned} S(\rho\|\xi) &= \frac{1}{\ln(2)} \sum_{1 \leq i,j \leq n} |\langle x_i, y_j \rangle|^2 \left(p_i \ln(p_i) - p_i \ln(q_j)\right) \\ &= \frac{1}{\ln(2)} \sum_{1 \leq i,j \leq n} |\langle x_i, y_j \rangle|^2 \left(q_j - p_i - \frac{1}{2}\eta''(\gamma_{ij})(p_i - q_j)^2\right) \end{aligned}$$

for some choice of real numbers $\{\gamma_{ij}\}$, where each γ_{ij} lies between p_i and q_j. In particular, this means that $0 < \gamma_{ij} \leq 1$, implying that $-\eta''(\gamma_{ij}) \geq 1$, for each choice of i and j. Consequently we have

$$S(\rho\|\xi) \geq \frac{1}{2\ln(2)} \sum_{1 \leq i,j \leq n} |\langle x_i, y_j \rangle|^2 (p_i - q_j)^2 = \frac{1}{2\ln(2)} \|\rho - \xi\|_2^2$$

as required. □

The following corollary represents a simple application of this fact. (We could just as easily prove it using analogous facts about the Shannon entropy, but the proof is essentially the same.)

Corollary 10.7. *Let $\mathcal{X}$ be a complex Euclidean space and let $n = \dim(\mathcal{X})$. It holds that $0 \leq S(\rho) \leq \log(n)$ for all $\rho \in \mathrm{D}(\mathcal{X})$. Furthermore, $\rho = \mathbb{1}/n$ is the unique density operator in $\mathrm{D}(\mathcal{X})$ having von Neumann entropy equal to $\log(n)$.*

Proof. The vector of eigenvalues $\lambda(\rho)$ of any density operator $\rho \in \mathrm{D}(\mathcal{X})$ is a probability vector, so $S(\rho) = H(\lambda(\rho))$ is a sum of nonnegative terms, which implies $S(\rho) \geq 0$. To prove the upper bound, let us assume ρ is a positive definite density operator, and consider the relative entropy $S(\rho\|\mathbb{1}/n)$. We have

$$0 \leq S(\rho\|\mathbb{1}/n) = -S(\rho) - \log(1/n)\,\mathrm{Tr}(\rho) = -S(\rho) + \log(n).$$

Therefore $S(\rho) \leq \log(n)$, and when ρ is not equal to $\mathbb{1}/n$ the inequality becomes strict. For density operators ρ that are not positive definite, the result follows from the continuity of von Neumann entropy. □

Now let us prove two simple properties of the von Neumann entropy: *subadditivity* and *concavity*. These properties also hold for the Shannon entropy—and while it is not difficult to prove them directly for the Shannon entropy, we get the properties for free once they are established for the von Neumann entropy.

When we refer to the von Neumann entropy of some collection of registers, we mean the von Neumann entropy of the state of those registers at some instant. For example, if X and Y are registers and $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ is the state of the pair (X, Y) at some instant, then

$$S(\mathsf{X}, \mathsf{Y}) = S(\rho), \quad S(\mathsf{X}) = S(\rho^{\mathsf{X}}), \quad \text{and} \quad S(\mathsf{Y}) = S(\rho^{\mathsf{Y}}),$$

where, in accordance with standard conventions, we have written $\rho^{\mathsf{X}} = \mathrm{Tr}_{\mathcal{Y}}(\rho)$ and $\rho^{\mathsf{Y}} = \mathrm{Tr}_{\mathcal{X}}(\rho)$. We often state properties of the von Neumann entropy in terms of registers, with the understanding that whatever statement is being discussed holds for all or some specified subset of the possible states of these registers. A similar convention is used for the Shannon entropy (for classical registers).

Theorem 10.8 (Subadditivity of von Neumann entropy)**.** *Let* X *and* Y *be quantum registers. For every state of the pair* (X, Y) *we have*

$$S(\mathsf{X}, \mathsf{Y}) \leq S(\mathsf{X}) + S(\mathsf{Y}).$$

Proof. Assume that the state of the pair (X, Y) is $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$. We will prove the theorem for ρ positive definite, from which the general case follows by continuity.

Consider the quantum relative entropy $S(\rho^{\mathsf{XY}} \| \rho^{\mathsf{X}} \otimes \rho^{\mathsf{Y}})$. Using the formula

$$\log(P \otimes Q) = \log(P) \otimes \mathbb{1} + \mathbb{1} \otimes \log(Q)$$

we find that

$$S(\rho^{\mathsf{XY}} \| \rho^{\mathsf{X}} \otimes \rho^{\mathsf{Y}}) = -S(\rho^{\mathsf{XY}}) + S(\rho^{\mathsf{X}}) + S(\rho^{\mathsf{Y}}).$$

By Theorem 10.6 we have $S(\rho^{\mathsf{XY}} \| \rho^{\mathsf{X}} \otimes \rho^{\mathsf{Y}}) \geq 0$, which completes the proof. □

In the next lecture we will prove a much stronger version of subadditivity, which is aptly named: *strong subadditivity*. It will imply the truth of the previous theorem, but it is instructive to compare the very easy proof above with the much more difficult proof of strong subadditivity.

Subadditivity also holds for the Shannon entropy:

$$H(\mathsf{X}, \mathsf{Y}) \leq H(\mathsf{X}) + H(\mathsf{Y})$$

for any choice of classical registers X and Y. This is simply a special case of the above theorem, where the density operator ρ is diagonal with respect to the standard basis of $\mathcal{X} \otimes \mathcal{Y}$.

Subadditivity implies that the von Neumann entropy is concave, as is established by the proof of the following theorem.

Theorem 10.9 (Concavity of von Neumann entropy)**.** *Let* $\rho, \xi \in \mathrm{D}(\mathcal{X})$ *and* $\lambda \in [0, 1]$*. It holds that*

$$S(\lambda\rho + (1-\lambda)\xi) \geq \lambda S(\rho) + (1-\lambda) S(\xi).$$

Proof. Let Y be a register corresponding to a single qubit, so that its associated space is $\mathcal{Y} = \mathbb{C}^{\{0,1\}}$. Consider the density operator

$$\sigma = \lambda\rho \otimes E_{0,0} + (1-\lambda)\xi \otimes E_{1,1},$$

and suppose that the state of the registers (X, Y) is described by σ. We have

$$S(\mathsf{X}, \mathsf{Y}) = \lambda S(\rho) + (1 - \lambda) S(\xi) + H(\lambda),$$

which is easily established by considering spectral decompositions of ρ and ξ. (Here we have referred to the *binary entropy function* $H(\lambda) = -\lambda \log(\lambda) - (1 - \lambda) \log(1 - \lambda)$.) Furthermore, we have

$$S(\mathsf{X}) = S(\lambda \rho + (1 - \lambda) \xi)$$

and

$$S(\mathsf{Y}) = H(\lambda).$$

It follows by subadditivity that

$$\lambda S(\rho) + (1 - \lambda) S(\xi) + H(\lambda) \leq S(\lambda \rho + (1 - \lambda) \xi) + H(\lambda)$$

which proves the theorem. □

Concavity also holds for the Shannon entropy as a simple consequence of this theorem, as we may take ρ and ξ to be diagonal with respect to the standard basis.

10.3 Conditional entropy and mutual information

Let us finish off the lecture by defining a few more quantities associated with the von Neumann entropy. We will not be able to say very much about these quantities until after we prove strong subadditivity in the next lecture.

Classically we define the *conditional Shannon entropy* as follows for two classical registers X and Y:

$$H(\mathsf{X}|\mathsf{Y}) = \sum_a \Pr[\mathsf{Y} = a] H(\mathsf{X}|\mathsf{Y} = a).$$

This quantity represents the expected value of the entropy of X given that you know the value of Y. It is not hard to prove that

$$H(\mathsf{X}|\mathsf{Y}) = H(\mathsf{X}, \mathsf{Y}) - H(\mathsf{Y}).$$

It follows from subadditivity that

$$H(\mathsf{X}|\mathsf{Y}) \leq H(\mathsf{X}).$$

The intuition is that your uncertainty can only increase when you know less.

In the quantum setting the first definition does not really make sense, so we use the second fact as our definition—the *conditional von Neumann entropy* of X given Y is

$$S(\mathsf{X}|\mathsf{Y}) = S(\mathsf{X}, \mathsf{Y}) - S(\mathsf{Y}).$$

Now we start to see some strangeness: we can have $S(\mathsf{Y}) > S(\mathsf{X}, \mathsf{Y})$, as we will if (X, Y) is in a pure, non-product state. This means that $S(\mathsf{X}|\mathsf{Y})$ can be negative, but such is life.

Next, the (classical) *mutual information* between two classical registers X and Y is defined as

$$I(\mathsf{X} : \mathsf{Y}) = H(\mathsf{X}) + H(\mathsf{Y}) - H(\mathsf{X}, \mathsf{Y}).$$

This can alternately be expressed as

$$I(\mathsf{X} : \mathsf{Y}) = H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X}) = H(\mathsf{X}) - H(\mathsf{X}|\mathsf{Y}).$$

We view this quantity as representing the amount of information in X about Y and vice versa. The *quantum mutual information* is defined similarly:

$$S(\mathsf{X} : \mathsf{Y}) = S(\mathsf{X}) + S(\mathsf{Y}) - S(\mathsf{X}, \mathsf{Y}).$$

At least we know from subadditivity that this quantity is always nonnegative. We will, however, need to further develop our understanding before we can safely associate any intuition with this quantity.