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

Lecture 12: Holevo's theorem and Nayak's bound

In this lecture we will prove Holevo's theorem. This is a famous theorem in quantum information theory, which is often informally summarized by a statement along the lines of this:

> *It is not possible to communicate more than n classical bits of information by the transmission of n qubits alone.*

Although this is an implication of Holevo's theorem, the theorem itself says more, and is stated in more precise terms that has no resemblance to the above statement. After stating and proving Holevo's theorem, we will discuss an interesting application of this theorem to the notion of a *quantum random access code*. In particular, we will prove Nayak's bound, which demonstrates that quantum random access codes are surprisingly limited in power.

12.1 Holevo's theorem

We will first discuss some of the concepts that Holevo's theorem concerns, and then state and prove the theorem itself. Although the theorem is difficult to prove from first principles, it turns out that there is a very simple proof that makes use of the strong subadditivity of the von Neumann entropy. Having proved strong subadditivity in the previous lecture, we will naturally opt for this simple proof.

12.1.1 Mutual information

Recall that if A and B are classical registers, whose values are distributed in some particular way, then the *mutual information* between A and B for this distribution is defined as

$$\begin{aligned} I(\mathsf{A}:\mathsf{B}) &\triangleq H(\mathsf{A}) + H(\mathsf{B}) - H(\mathsf{A},\mathsf{B}) \\ &= H(\mathsf{A}) - H(\mathsf{A}|\mathsf{B}) \\ &= H(\mathsf{B}) - H(\mathsf{B}|\mathsf{A}). \end{aligned}$$

The usual interpretation of this quantity is that it describes how many bits of information about B are, on average, revealed by the value of A; or equivalently, given that the quantity is symmetric in A and B, how many bits of information about A are revealed by the value of B. Like all quantities involving the Shannon entropy, this interpretation should be understood to really only be meaningful in an asymptotic sense.

To illustrate the intuition behind this interpretation, let us suppose that A and B are distributed in some particular way, and Alice looks at the value of A. As Bob does not know what value Alice sees, he has $H(\mathsf{A})$ bits of uncertainty about her value. After sampling B, Bob's average uncertainty about Alice's value becomes $H(\mathsf{A}|\mathsf{B})$, which is always at most $H(\mathsf{A})$ and is less assuming that A and B are correlated. Therefore, by sampling B, Bob has decreased his uncertainty of Alice's value by $I(\mathsf{A}:\mathsf{B})$ bits.

In analogy to the above formula we have defined the *quantum mutual information* between two registers X and Y as

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

Although Holevo's theorem does not directly concern the quantum mutual information, it is nevertheless related and indirectly appears in the proof.

12.1.2 Accessible information

Imagine that Alice wants to communicate classical information to Bob. In particular, suppose Alice wishes to communicate to Bob information about the value of a classical register A, whose possible values are drawn from some set Σ and where $p \in \mathbb{R}^{\Sigma}$ is the probability vector that describes the distribution of these values:

$$p(a) = \Pr[\mathsf{A} = a]$$

for each $a \in \Sigma$.

The way that Alice chooses to do this is by preparing a quantum register X in some way, depending on A, after which X is sent to Bob. Specifically, let us suppose that $\{\rho_a : a \in \Sigma\}$ is a collection of density operators in $\mathrm{D}(\mathcal{X})$, and that Alice prepares X in the state ρ_a for whichever $a \in \Sigma$ is the value of A. The register X is sent to Bob, and Bob measures it to gain information about the value of Alice's register A.

One possible approach that Bob could take would be to measure X with respect to some measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, chosen so as to maximize the probability that his measurement result b agrees with Alice's sample a (as was discussed in Lecture 8). We will not, however, make the assumption that this is Bob's approach, and in fact we will not even assume that Bob chooses a measurement whose outcomes agree with Σ. Instead, we will consider a completely general situation in which Bob chooses a measurement

$$\mu : \Gamma \to \mathrm{Pos}(\mathcal{X}) : b \mapsto P_b$$

with which to measure the register X sent by Alice. Let us denote by B a classical register that stores the result of this measurement, so that the pair of registers (A, B) is then distributed as follows:

$$\Pr[(\mathsf{A}, \mathsf{B}) = (a, b)] = p(a) \langle P_b, \rho_a \rangle$$

for each $(a, b) \in \Sigma \times \Gamma$. The amount of information that Bob gains about A by means of this process is given by the mutual information $I(\mathsf{A} : \mathsf{B})$.

The *accessible information* is the maximum value of $I(\mathsf{A} : \mathsf{B})$ that can be achieved by Bob, over all possible measurements. More precisely, the accessible information of the *ensemble*

$$\mathcal{E} = \{(p(a), \rho_a) : a \in \Sigma\}$$

is defined as the maximum of $I(\mathsf{A} : \mathsf{B})$ over all possible choices of the set Γ and the measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{X})$, assuming that the pair (A, B) is distributed as described above for this choice of a measurement. (Given that there is no upper bound on the size of Bob's outcome set Γ, it is not obvious that the accessible information of a given ensemble $\mathcal{E}$ is always achievable for some fixed choice of a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{X})$. It turns out, however, that there is always an achievable maximum value for $I(\mathsf{A} : \mathsf{B})$ that Bob reaches when his set of outcomes Γ has size at most $\dim(\mathcal{X})^2$.) We will write $I_{\mathrm{acc}}(\mathcal{E})$ to denote the accessible information of the ensemble $\mathcal{E}$.

12.1.3 The Holevo quantity

The last quantity that we need to discuss before stating and proving Holevo's theorem is the *Holevo χ-quantity*. Let us consider again an ensemble $\mathcal{E} = \{(p(a), \rho_a) : a \in \Sigma\}$, where each ρ_a is a density operator on $\mathcal{X}$ and $p \in \mathbb{R}^{\Sigma}$ is a probability vector. For such an ensemble we define the Holevo χ-quantity of $\mathcal{E}$ as

$$\chi(\mathcal{E}) \triangleq S\left(\sum_{a\in\Sigma} p(a)\rho_a\right) - \sum_{a\in\Sigma} p(a)S(\rho_a).$$

Notice that the quantity $\chi(\mathcal{E})$ is always nonnegative, which follows from the concavity of the von Neumann entropy.

One way to think about this quantity is as follows. Consider the situation above where Alice has prepared the register X depending on the value of A, and Bob has received (but not yet measured) the register X. From Bob's point of view, the state of X is therefore

$$\rho = \sum_{a\in\Sigma} p(a)\rho_a.$$

If, however, Bob were to learn that the value of A is $a \in \Sigma$, he would then describe the state of X as ρ_a. The quantity $\chi(\mathcal{E})$ therefore represents the average decrease in the von Neumann entropy of X that Bob would expect from learning the value of A.

Another way to view the quantity $\chi(\mathcal{E})$ is to consider the state of the pair (A, X) in the situation just considered, which is

$$\xi = \sum_{a\in\Sigma} p(a)E_{a,a} \otimes \rho_a.$$

We have

$$\begin{aligned} S(\mathsf{A}, \mathsf{X}) &= H(p) + \sum_{a\in\Sigma} p(a)S(\rho_a), \\ S(\mathsf{A}) &= H(p), \\ S(\mathsf{X}) &= S\left(\sum_{a\in\Sigma} p(a)\rho_a\right), \end{aligned}$$

and therefore $\chi(\mathcal{E}) = S(\mathsf{A}) + S(\mathsf{X}) - S(\mathsf{A}, \mathsf{X}) = S(\mathsf{A} : \mathsf{X})$.

12.1.4 Statement and proof of Holevo's theorem

Now we are prepared to state and prove Holevo's theorem. The formal statement of the theorem follows.

Theorem 12.1 (Holevo's theorem)**.** *Let $\mathcal{E} = \{(p(a), \rho_a) : a \in \Sigma\}$ be an ensemble of density operators over some complex Euclidean space $\mathcal{X}$. It holds that $I_{acc}(\mathcal{E}) \leq \chi(\mathcal{E})$.*

Proof. Suppose Γ is a finite, nonempty set and $\mu : \Gamma \to \mathrm{Pos}(\mathcal{X}) : b \mapsto P_b$ is a measurement on $\mathcal{X}$. Let $\mathcal{A} = \mathbb{C}^{\Sigma}$, $\mathcal{B} = \mathbb{C}^{\Gamma}$, and let us regard μ as a channel of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{B})$ defined as

$$\Phi(X) = \sum_{b\in\Gamma} \langle P_b, X\rangle E_{b,b}$$

for each $X \in \mathrm{L}(\mathcal{X})$. Like every channel, there exists a Stinespring representation for Φ:

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(VXV^*)$$

for some choice of a complex Euclidean space $\mathcal{Z}$ and a linear isometry $V \in \mathrm{U}(\mathcal{X}, \mathcal{B} \otimes \mathcal{Z})$.

Now define two density operators, $\sigma \in \mathrm{D}(\mathcal{A} \otimes \mathcal{X})$ and $\xi \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B} \otimes \mathcal{Z})$, as follows:

$$\sigma = \sum_{a \in \Sigma} p(a) E_{a,a} \otimes \rho_a \qquad \text{and} \qquad \xi = (\mathbb{1}_{\mathcal{A}} \otimes V)\sigma(\mathbb{1}_{\mathcal{A}} \otimes V)^*.$$

Given that V is an isometry, the following equalities hold:

$$\begin{aligned} S(\xi^{\mathsf{A}}) &= S(\sigma^{\mathsf{A}}) = H(p) \\ S(\xi^{\mathsf{ABZ}}) &= S(\sigma^{\mathsf{AX}}) = H(p) + \sum_{a \in \Sigma} p(a) S(\rho_a) \\ S(\xi^{\mathsf{BZ}}) &= S(\sigma^{\mathsf{X}}) = S\left(\sum_{a \in \Sigma} p(a) \rho_a \right), \end{aligned}$$

and therefore, for the state $\xi \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B} \otimes \mathcal{Z})$, we have

$$S(\mathsf{A} : \mathsf{B}, \mathsf{Z}) = S(\mathsf{A}) + S(\mathsf{B}, \mathsf{Z}) - S(\mathsf{A}, \mathsf{B}, \mathsf{Z}) = S\left(\sum_{a \in \Sigma} p(a) \rho_a \right) - \sum_{a \in \Sigma} p(a) S(\rho_a) = \chi(\mathcal{E}).$$

By the strong subadditivity of the von Neumann entropy, we have

$$S(\mathsf{A} : \mathsf{B}) \leq S(\mathsf{A} : \mathsf{B}, \mathsf{Z}) = \chi(\mathcal{E}).$$

Noting that the state $\xi^{\mathsf{AB}} \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B})$ takes the form

$$\xi = \sum_{a \in \Sigma} \sum_{b \in \Gamma} p(a) \langle P_b, \rho_a \rangle E_{a,a} \otimes E_{b,b},$$

we see that the quantity $S(\mathsf{A} : \mathsf{B})$ is equal to the accessible information $I_{\mathrm{acc}}(\mathcal{E})$ for an optimally chosen measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{X})$. It follows that $I_{\mathrm{acc}}(\mathcal{E}) \leq \chi(\mathcal{E})$ as required.

□

As discussed at the beginning of the lecture, this theorem implies that Alice can communicate no more than n classical bits of information to Bob by sending n qubits alone. If the register X comprises n qubits, and therefore $\mathcal{X}$ has dimension 2^n, then for any ensemble

$$\mathcal{E} = \{(p(a), \rho_a) \,:\, a \in \Sigma\}$$

of density operators on $\mathcal{X}$ we have

$$\chi(\mathcal{E}) \leq S\left(\sum_{a \in \Sigma} p(a) \rho_a \right) \leq n.$$

This means that for any choice of the register A, the ensemble $\mathcal{E}$, and the measurement that determines the value of a classical register B, we must have $I(\mathsf{A} : \mathsf{B}) \leq n$. In other words, Bob can learn no more than n bits of information by means of the process he and Alice have performed.

12.2 Nayak's bound

We will now consider a related, but nevertheless different setting from the one that Holevo's theorem concerns. Suppose now that Alice has m bits, and she wants to encode them into fewer than n qubits in such a way that Bob can recover not the entire string of bits, but rather any *single* bit (or small number of bits) of his choice. Given that Bob will only learn a very small amount of information by means of this process, the possibility that Alice could do this does not violate Holevo's theorem in any obvious way.

This idea has been described as a "quantum phone book." Imagine a very compact phone book implemented using quantum information. The user measures the qubits forming the phone book using a measurement that is unique to the individual whose number is being sought. The user's measurement destroys the phone book, so only a small number of digits of information are effectively transmitted by sending the phone book. Perhaps it is not unreasonable to hope that such a phone book containing 100,000 numbers could be constructed using, say, 1,000 qubits?

Here is an example showing that something nontrivial along these lines can be realized. Suppose Alice wants to encode 2 bits $a, b \in \{0,1\}$ into one qubit so that when she sends this qubit to Bob, he can pick a two-outcome measurement giving him either a or b with reasonable probability. Define

$$|\psi(\theta)\rangle = \cos(\theta)\,|0\rangle + \sin(\theta)\,|1\rangle$$

for $\theta \in [0, 2\pi]$. Alice encodes ab as follows:

$$\begin{aligned} 00 &\rightarrow |\psi(\pi/8)\rangle \\ 10 &\rightarrow |\psi(3\pi/8)\rangle \\ 11 &\rightarrow |\psi(5\pi/8)\rangle \\ 01 &\rightarrow |\psi(7\pi/8)\rangle \end{aligned}$$

Alice sends the qubit to Bob. If Bob wants to decode a, be measures in the $\{|0\rangle, |1\rangle\}$ basis, and if he wants to decode b he measures in the $\{|+\rangle, |-\rangle\}$ basis (where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$). A simple calculation shows that Bob will correctly decode the bit he has chosen with probability $\cos^2(\pi/8) \approx .85$. There does not exist an analogous classical scheme that allow Bob to do better than randomly guessing for at least one of his two possible choices.

12.2.1 Definition of quantum random access encodings

In more generality, we define a *quantum random access encoding* according to the definition that follows. Here, and for the remainder of the lecture, we let $\Sigma = \{0,1\}$.

Definition 12.2. Let m and n be positive integers, and let $p \in [0,1]$. An $m \overset{p}{\mapsto} n$ quantum random access encoding is a function

$$R : \Sigma^m \rightarrow \mathrm{D}\left(\mathbb{C}^{\Sigma^n}\right) : a_1 \cdots a_m \mapsto \rho_{a_1 \cdots a_m}$$

such that the following holds. For each $j \in \{1, \ldots, m\}$ there exists a measurement

$$\left\{P_0^j, P_1^j\right\} \subset \mathrm{Pos}\left(\mathbb{C}^{\Sigma^n}\right)$$

such that

$$\left\langle P_{a_j}^j, \rho_{a_1 \cdots a_m} \right\rangle \geq p$$

for every $j \in \{1, \ldots, m\}$ and every choice of $a_1 \cdots a_m \in \Sigma^m$.

For example, the above example is a $2 \overset{.85}{\mapsto} 1$ quantum random access encoding.

12.2.2 Fano's inequality

In order to determine whether $m \overset{p}{\mapsto} n$ quantum random access codes exist for various choices of the parameters n, m, and p, we will need a result from classical information theory known as *Fano's inequality*. When considering this result, recall that the binary entropy function is defined as

$$H(\lambda) = -\lambda \log(\lambda) - (1-\lambda)\log(1-\lambda)$$

for $\lambda \in [0,1]$.

Theorem 12.3 (Fano's inequality)**.** *Let* A *and* B *be classical registers taking values in some finite set* Γ *and let* $q = \Pr[\mathsf{A} \neq \mathsf{B}]$*. It holds that*

$$H(\mathsf{A}|\mathsf{B}) \leq H(q) + q\log(|\Gamma| - 1).$$

Proof. Define a new register C whose value is determined by A and B as follows:

$$\mathsf{C} = \begin{cases} 1 & \text{if } \mathsf{A} \neq \mathsf{B} \\ 0 & \text{if } \mathsf{A} = \mathsf{B}. \end{cases}$$

Let us first note that

$$H(\mathsf{A}|\mathsf{B}) = H(\mathsf{C}|\mathsf{B}) + H(\mathsf{A}|\mathsf{B},\mathsf{C}) - H(\mathsf{C}|\mathsf{A},\mathsf{B}).$$

This holds for any choice of registers as a result of the following equations:

$$\begin{aligned} H(\mathsf{A}|\mathsf{B}) &= H(\mathsf{A},\mathsf{B}) - H(\mathsf{B}), \\ H(\mathsf{C}|\mathsf{B}) &= H(\mathsf{B},\mathsf{C}) - H(\mathsf{B}), \\ H(\mathsf{A}|\mathsf{B},\mathsf{C}) &= H(\mathsf{A},\mathsf{B},\mathsf{C}) - H(\mathsf{B},\mathsf{C}), \\ H(\mathsf{C}|\mathsf{A},\mathsf{B}) &= H(\mathsf{A},\mathsf{B},\mathsf{C}) - H(\mathsf{A},\mathsf{B}). \end{aligned}$$

Next, note that

$$H(\mathsf{C}|\mathsf{B}) \leq H(\mathsf{C}) = H(q).$$

Finally, we have $H(\mathsf{C}|\mathsf{A},\mathsf{B}) = 0$ because C is determined by A and B. So, at this point we conclude

$$H(\mathsf{A}|\mathsf{B}) \leq H(q) + H(\mathsf{A}|\mathsf{B},\mathsf{C}).$$

It remains to put an upper bound on $H(\mathsf{A}|\mathsf{B},\mathsf{C})$. We have

$$H(\mathsf{A}|\mathsf{B},\mathsf{C}) = \Pr[\mathsf{C}=0]H(\mathsf{A}|\mathsf{B},\mathsf{C}=0) + \Pr[\mathsf{C}=1]H(\mathsf{A}|\mathsf{B},\mathsf{C}=1).$$

We also have

$$H(\mathsf{A}|\mathsf{B},\mathsf{C}=0) = 0,$$

because $\mathsf{C} = 0$ implies $\mathsf{A} = \mathsf{B}$, and

$$H(\mathsf{A}|\mathsf{B},\mathsf{C}=1) \leq \log(|\Gamma| - 1)$$

because $\mathsf{C} = 1$ implies that $\mathsf{A} \neq \mathsf{B}$, so the largest the uncertainty of A given $\mathsf{B} = b$ can be is $\log(|\Gamma| - 1)$, which is the case when A is uniform over all elements of Γ besides b. Thus

$$H(\mathsf{A}|\mathsf{B}) \leq H(q) + q\log(|\Gamma| - 1)$$

as required. □

The following special case of Fano's inequality, where $\Gamma = \{0,1\}$ and A is uniformly distributed, will be useful.

Corollary 12.4. *Let A be a uniformly distributed Boolean register and let B be any Boolean register. For $q = \Pr(\mathsf{A} = \mathsf{B})$ we have $I(\mathsf{A} : \mathsf{B}) \geq 1 - H(q)$.*

12.2.3 Statement and proof of Nayak's bound

We are now ready to state Nayak's bound, which implies that the quest for a compact quantum phone book was doomed to fail: any $m \overset{p}{\mapsto} n$ quantum random access code requires that n and m are linearly related, with the constant of proportionality tending to 1 as p approaches 1.

Theorem 12.5 (Nayak's bound)**.** *Let n and m be positive integers and let $p \in [1/2, 1]$. If there exists a $m \overset{p}{\mapsto} n$ quantum random access encoding, then $n \geq (1 - H(p))m$.*

To prove this theorem, we will first require the following lemma, which is a consequence of Holevo's theorem and Fano's inequality.

Lemma 12.6. *Suppose $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ are density operators, $\{Q_0, Q_1\} \subseteq \mathrm{Pos}(\mathcal{X})$ is a measurement, and $q \in [1/2, 1]$. If it is the case that*

$$\langle Q_b, \rho_b \rangle \geq q$$

for $b \in \Sigma$, then

$$S\left(\frac{\rho_0 + \rho_1}{2}\right) - \frac{S(\rho_0) + S(\rho_1)}{2} \geq 1 - H(q)$$

Proof. Let A and B be classical Boolean registers, let $p \in \mathbb{R}^{\Sigma}$ be the uniform probability vector (meaning $p(0) = p(1) = 1/2$), and assume that

$$\Pr[(\mathsf{A}, \mathsf{B}) = (a, b)] = p(a) \langle Q_b, \rho_a \rangle$$

for $a, b \in \Sigma$. By Holevo's theorem we have

$$I(\mathsf{A} : \mathsf{B}) \leq S\left(\frac{\rho_0 + \rho_1}{2}\right) - \frac{S(\rho_0) + S(\rho_1)}{2},$$

and by Fano's inequality we have

$$I(\mathsf{A} : \mathsf{B}) \geq 1 - H(\Pr[\mathsf{A} = \mathsf{B}])) \geq 1 - H(q),$$

from which the lemma follows. □

Proof of Theorem 12.5. Suppose we have some $m \overset{p}{\mapsto} n$ quantum random access encoding

$$R : a_1 \cdots a_m \mapsto \rho_{a_1 \cdots a_m}.$$

For $0 \leq k \leq m - 1$ let

$$\rho_{a_1 \cdots a_k} = \frac{1}{2^{m-k}} \sum_{a_{k+1} \cdots a_m \in \Sigma^{m-k}} \rho_{a_1 \cdots a_m},$$

and note that

$$\rho_{a_1 \cdots a_k} = \frac{1}{2}(\rho_{a_1 \cdots a_k 0} + \rho_{a_1 \cdots a_k 1}).$$

By the assumption that R is a random access encoding, we have that there exists a measurement $\{P_0^k, P_1^k\}$, for $1 \leq k \leq m$, that satisfies

$$\left\langle P_b^k, \rho_{a_1 \cdots a_{k-1} b} \right\rangle \geq p$$

for each $b \in \Sigma$. Thus, by Lemma 12.6,

$$S(\rho_{a_1 \cdots a_{k-1}}) \geq \frac{1}{2}(S(\rho_{a_1 \cdots a_{k-1} 0}) + S(\rho_{a_1 \cdots a_{k-1} 1})) + (1 - H(p))$$

for $1 \leq k \leq m$ and all choices of $a_1 \cdots a_{k-1}$. By applying this inequality repeatedly, we conclude that

$$m(1 - H(p)) \leq S(\rho) \leq n,$$

which completes the proof. □

It can be shown that there exists a classical random access encoding $m \overset{p}{\mapsto} n$ for any $p > 1/2$ provided that $n \in (1 - H(p))m + O(\log m)$. Thus, asymptotically speaking, there is no significant advantage to be gained from quantum random access codes over classical random access codes.