Introduction to quantum computation: Quantum Communication

Lecture Notes for Quantum Computing

Lecturer: Dorit Aharonov   Scribe: Sharon Pecker

June 13, 2001

Summary of the thirteenth week of semester 2 (13.6.2001)

1 Definitions

1.1 Classical Entropy Of A Single Random Variable

We define the entropy $H(X)$ of a random variable X with a distribution of $\{P(x_n)\}$ as:

$$H(X) := -\sum P(x_n)\log(P(x_n))$$

We see that for a constant random variable there is no information as we might have excpected, as $H(X) = 0$.

Later we will see that for a uniformly distributed random variable we get the maximal value of information.

A way for us to understand the meaning of our definition could be the following. Given a random variable, say X, with a known distribution $\{P(x_i)\}$. Given that we need to find a way to encode the values of X so that we would be able to send them as strings of bits. $H(X)$ would be the minimal average length of string we would be required to send. This minimal average length can be reached, as we know, using the algorithm of Hoffman.

For example if we have the random variable X that gets the value 1 in a probability of $\frac{1}{2}$, and 2 or 3, both in a probability of $\frac{1}{4}$. Hoffman's algorithm would give the best code, that would use one bit for the first value and two for the others. This would give an average of $1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} = \frac{3}{2}$ which is exactly $H(X)$. So H does represent in a good way the information existing in a random variable.

Let us examine two properties of H (Proofs can be seen in the book of Nielsen and Chuang):

1. **Convexity**: for any two random variables X and Y and a number q *s.t.* $0 \le q \le 1$ we get

$$H(qX + (1-q)Y) \ge qH(X) + (1-q)H(Y)$$

2. **H gets its maximal value for a uniform random variable**: We get this from the last property as every random variable can be described as a sum of the constant random variables whose entropy is zero. Thus, we get for a random variable with N different values:

$$H(X) \le H(\bar{X}) = \log N$$

For example, for a boolean random variable X with a probability of p to get the value of 1 we get the relation between $H(X)$ and p that is shown in the graph below:

HERE SHOULD BE THE GRAPH

For a random variable that is string $s \in \{0,1\}^n$, $H(s) \le n$

1.2 Classical Entropy Of Dependent Random Variables

Let us continue with defining entropy of a compound random variable as:

$$H(X,Y) := -\sum P_{(X,Y)} \log \left(P_{(X,Y)}\right)$$

Conditional entropy would be defined as:

$$H(X|Y) := -\sum P_y H(X|Y=y) = H(X,Y) - H(Y)$$

And mutual entropy is defined as:

$$H(X:Y) := H(X) + H(Y) - H(X,Y) = H(Y:X)$$

We can derive the following properties of the above stated definitions, given two random varaibles X and Y:

1. For X, Y that are equal (or that are fully dependent of eachother) we get:
$$H(X,Y) = H(X)$$
$$H(X|Y) = 0$$
$$H(X:Y) = H(X)$$

2. For X, Y that are independent we get:
$$H(X,Y) = H(X) + H(Y)$$
$$H(X|Y) = H(X)$$
$$H(X:Y) = 0$$

3. The following inequation is always true:
$$0 \leq H(X|Y) \leq H(X) \leq H(X,Y) \leq H(X) + H(Y)$$
Moreover, the second and last equations are attained only for independent variables. The first one, only for fully dependent variables.

4. $$H(X:Y) \leq H(X:YZ)$$

5. For $f(Y)$, a function of the random variable Y we get:
$$H(X:f(Y) \leq H(X:Y)$$
Equation is attained iff f is one to one.

1.3 Entropy Of Quantum Random Variables

When trying to define similar definitions for quantum variables, we expect them to lead to conclusions similar to those we have just reached a section ago. For one, we excpect a constant pure state to have no information "contained" in it, for it can always be found a base, in which measuring the state would give a constant result. Let us try the following definition.

For a variable given as a density matrix ρ we define the entropy as:
$$S(\rho) := -tr(\rho \log \rho)$$

The definitions of compound, conditional and mutual entropies follow as in the classical case given the appropriate density matrices and indeed, for a (constant) pure state X, $S(\rho_X) = 0$

Again, several properties for any density matrix ρ:

1. **Convexity**
2. **Qubits**: For ρ a density matrix of a set of n qubits:
$$S(\rho) \leq n$$
3. $$S(\rho) \geq 0$$
4. $$S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$$
5. **For a unitary matrix U**:
$$S(U\rho U^{\dagger}) = S(\rho)$$
Thus, given a density matrix with eigenvalues λ_i, we get $S(\rho) = -\sum \lambda_i \log(\lambda_i)$, for we can put its eigenvalues on its diagonal using a unitary matrix.
6. $$S(A|BC) \leq S(A|B)$$
Therefore also:
$$S(A : B) \leq S(A : BC)$$
and
$$S(A : f(B)) \leq S(A : B)$$
as in the classical case.

All of that might lead us to think that all that was true for classical entropy is true for quantum entropy, but this is not the case, for $S(A) \leq S(X, Y)$ is not always true!

This means we can find a state s.t. measuring would increase its entropy. For example, let us examine an EPR pair. $S(A, B)$ would result to be zero as the state is constant. However, after measuring, say, the second qubit, we get S(A)=1.

This leads to the definition of the entanglement:

$$E(A : B) := S(A)$$

2 Holavo Bound

2.1 Introduction

Say that Alice holds a random variable X (s.t. $H(X) = n$) that she wants to send to Bob using the minimal number of qubits. She uses a code $|\alpha_x\rangle$ in the following manner. According to the distribution of X, $\{P_x\}$, she sends $\rho = \sum P_x |\alpha_x\rangle\langle\alpha_x|$.

Bob is then allowed to perform any unitary operation and measure the outcome. Say he gets a random variable Y. As close as it gets to X, $H(X:Y)$ gets larger. We have said earlier that the minimal number of bits needed for sending ρ is $S(\rho)$, so as Alice wants to use the minimal number of bits m, she actually wants $S(\rho)$ to be minimal. But as we will soon see, the Holavo Bound says that:

$$H(X:Y) \leq S(\rho)$$

This means that in order for $H(X:Y)$ to reach its maximal value, $H(X)$, Alice needs to use at least $m \leq n$ qubits.

2.2 Proof

Alice has X as $\sum P_x |x\rangle\langle x|_A$ and her coding ρ. We denote the bits used for coding with a C . Alice can also create

$$\sum_x P_x |x\rangle\langle x|_A \otimes |\alpha_x\rangle\langle\alpha_x|_C \otimes |0\rangle\langle 0|_B$$

where Bob has the bit set denoted with a B.

Bob gets the C bits and is allowed to do whatever he wishes, which basically means he can perform a unitary transformation and measure B to get the random variable Y.

Lemma 1 *Let us denote the states of A and B after the operations of Bob as A' and B'. Then we get that*

$$S(A':B') \leq S(A:C)$$

Proof *First of all, $A = A'$ as bob has no access to A at all. Then, using the properties of S we stated above we get*

$$S(A':B') \leq S(A:B'C')$$

which would have been true for any C. Now, as unitary transformations does not change the entropy, the measurement of Bob leads to

$$S(A : B'C') \leq S(A : BC')$$

but the original B is a set of zeros, thus

$$S(A : BC') = S(A : C')$$

and this finishes the proof as it all sums to

$$S(A' : B') \leq S(A : C)$$

Now we see that this is exactly the Holavo bound as

$$S(A' : B') = S(A') + S(B') - S(A'B') = H(X) + H(Y) - H(XY) = H(X : Y)$$

and

$$S(A : C) = S(A) + S(C) - S(AC) = H(X) + S(\rho) - H(X) = S(\rho)$$

while $S(AC) = H(X)$ is true as $\sum P_x|x\rangle$ and $\sum P_x|x\rangle_A \otimes |\alpha_x\rangle_C$ are equivalent for the sake representing the random variable X, for $|x\rangle \otimes |\alpha_x\rangle$ are orthogonal.

We have reached the sad conclusion that quantum communication does not allow us use less bits than classical communication would require. This turns out to be true even if we allow a measurement of only one bit of the given strings. A measurement that would destroy all the other bits.

What advantage, then, quantum communication shows over classical communication?

3 The Problem Of Disjointness

3.1 Problem definition

Say that Alice has a bit string $X = \{x_i\}$ and Bob has a bit string $Y = \{y_i\}$. The objective is to find whether

$$Disj(X, Y) := \bigvee_{i=1}^{n} (x_i \wedge y_i)$$

without sending the strings. This can be understood also as a problem of finding a common element i of two groups X, Y.

3.2 Quantum solution

This problem reminds us of the Grover searching algorithm, given an oracle that would answer $z_i = x_i \wedge y_i$. *The gover algorithm requires* $O(\sqrt{n})$ *calls for the oracle.*

Now all we have left to do is define the oracle and find its complexity. Alice creates the $\log n + 3$ *bits state*

$$\sum_{i=1}^{n} |i\rangle|0\rangle|0\rangle|0\rangle$$

and than sets the first zero bit to x_i. *Thus, she can send Bob*

$$\sum_{i=1}^{n} |i\rangle|x_i\rangle|0\rangle|0\rangle$$

Now, Bob sets the second zero bit to y_i. *Thus, he can compute*

$$\sum_{i=1}^{n} |i\rangle|x_i\rangle|y_i\rangle|z_i\rangle$$

Now he can erase the $|y_i\rangle$ *bit and send the bits back to Alice, who can then erase her bit and result with*

$$\sum_{i=1}^{n} |i\rangle|z_i\rangle \otimes |0\rangle|0\rangle$$

which should be the answer of the oracle. All that required $O(\log n)$ *bits for each call for the oracle. It all sums to* $O(\sqrt{n}\log n)$, *where classically it would require* $\Theta(n)$ *cbits.*

This problem showed a quadratic improvement using quantum communication.

4 The Problem Of Equality

4.1 Problem definition

We have just seen a quadratic improvement for a solution using quantum communication. But we know of a problem whose solution can be exponentially improved using quantum computation. That is the problem of Deutsch *and Jozsa. To transform it into a communication problem let assume Alice has a string X and Bob has a string Y. Now we define the problem to be to find out whether $X = Y$ or $\Delta(X,Y) := \sum x_i \oplus y_i = \frac{N}{2}$ when we are promised that one of the above is always true.*

4.2 Quantum solution

To convert the problem above to Deutch-Jozsa we need to form a function that is either constant, or it gets zero for exactly half of the inputs. This would simply be $f(i) := x_i \oplus y_i$, since X and Y are either equal (thus, $f(i) \equiv 0$), or have exactly $\frac{N}{2}$ equal bits and $\frac{N}{2}$ nonequal bits (thus, $f(i) = 0$ for exactly half of the values of i).

We can use the same oracle as with the Disj problem, and since Deutch-Jozsa requires one or two calls for the oracle, we get a complexity of $O(\log n)$ qubits.

4.3 An upper bound for the classical solution

We will now see that a classical solution would require $O(n)$ bits. X and Y are described using n bits each, so we could draw an $(2^n \times 2^n)$ matrix that would contain 1's on its diagonal, as there $X = Y$ and 0's wherever $\Delta(X,Y) = 0$.

On each round Alice and Bob send one bit, one at a time. Let us define that for Alice 1 means left and 0 means right. Whereas for Bob, it would be up and down. After $\frac{n}{100}$ rounds of doing this we would be bounded inside a certain rectangle, one of $2^{\frac{n}{100}}$, which can be achieved according to the definitions of Alice's (and Bob's) separations between left and right (or up and down).

It should be clear that the goal of this protocol is to have all rectangles monochromatic. i.e., that no rectangle would suffer from having both a 1 and a 0 inside, for in that case Alice's answer would not be well defined. Also, we should notice, that since there are 2^n on the diagonal of the matrix, there is a rectangle larger than $(\frac{2^n}{2^{\frac{n}{100}}})^2$. That is so because some rectangle should have at least $\frac{2^n}{2^{\frac{n}{100}}}$ 1's in it.

There is, however, a theorem that states that there can be no rectangle of that size with no 0's in it. So, our classical protocol is useless as it would require $O(n)$. We've proved the existance of an exponential gap.

Still, we should not give up on classical protocols, as if a certain amount *of error is allowed, the classical-quantum gap vanishes, for both protocols give* $O(\log n)$.

5 A Major Quantum Gap

There already is a known problem such that there is an exponential gap between the classical and quantum solutions, even if we do allow a certain amount of error $\epsilon > 0$*, where* X *is a description of a state* $|\alpha\rangle$ *and* Y *is of a unitary matrix* U*. We will leave this without a proof.*