Lecture 21: Quantum communication complexity

April 6, 2006

In this lecture we will discuss how quantum information can allow for a significant reduction in the communication costs associated with various distributed tasks. Before we discuss how this is possible, however, it is important to mention that quantum information cannot reduce the cost of direct communication. This is due to Holevo's Theorem.

Holevo's Theorem

Let us begin with a simplified version of the theorem, which is sufficient to give the spirit of Holevo's Theorem, but not too much more.

Holevo's Theorem (informal, simplified version). *If Alice and Bob share no prior entanglement, then it is necessary for them to exchange at least n qubits (or bits) in order for Alice to communicate n classical bits of information to Bob. If they do share prior entanglement, then they must exchange at least $n/2$ qubits.*

I would like to mention the formal version of Holevo's Theorem, even though we have not covered the necessary concepts to understand it properly. The reason is that anyone who plans to continue studying quantum information after this course should not be under the impression that Holevo's Theorem is as simple as the statement above: it is both qualitatively and quantitatively deeper than the simplified version suggests.

Holevo's Theorem (formal version). *Let Σ and Γ be finite, nonempty sets and let $\mathcal{X}$ be the space corresponding to some quantum system. Let $\{\rho_a : a \in \Sigma\} \subset \mathrm{D}(\mathcal{X})$ be a collection of density operators, let $\{M_b : b \in \Gamma\}$ be a POVM on $\mathcal{X}$, and let $p \in \mathbb{R}(\Sigma)$ be a probability vector. Let A and B be random variables taking values in Σ and Γ, respectively, such that $\Pr[A = a] = p[a]$ and $\Pr[B = b \,|\, A = a] = \mathrm{Tr}(M_b \rho_a)$. Then*

$$I(A : B) \leq S\left(\sum_{a \in \Sigma} p[a]\rho_a\right) - \sum_{a \in \Sigma} p[a] S(\rho_a).$$

In this theorem, $I(A : B)$ is the *mutual information* between the random variables A and B, and $S(\cdot)$ is the *von Neumann entropy* function. We may interpret the statement of the theorem as follows. The random variable A, which has associated probability vector p, determines the "symbol" a (or string a if you prefer) that Alice wishes to send to Bob. Alice samples A to find some a, prepares a quantum state ρ_a, and sends it to Bob. Bob then measures this state with respect to some POVM $\{M_b : b \in \Gamma\}$. The random variable B describes the outcome of this measurement. You could imagine that $\Gamma = \Sigma$, and Bob is hoping that $\Pr[A = B]$ is large in order

to learn a, but the theorem is more general. The theorem implies that the amount of information that B contains about A (which is $I(A:B)$) is at most the quantity on the right-hand-side of the inequality. It is a simple property of the von Neumann entropy that this quantity, sometimes called the *Holevo quantity* or *Holevo χ quantity*, cannot be larger than the logarithm of the dimension of $\mathcal{X}$. This puts a lower bound on the size of $\mathcal{X}$ required for any given amount of information to have been transmitted from Alice to Bob.

Holevo's Theorem is easily proved once a fundamental property of the von Neumann entropy, known as *strong subadditivity*, is proved. Proving strong subadditivity, on the other hand, is highly nontrivial but not beyond your reach—for instance it takes about two lectures to prove in my advanced quantum information theory graduate course.

Communication complexity

Although quantum information cannot reduce the direct cost of communication as Holevo's Theorem implies, there are many distributed computational tasks that require communication but not direct communication in an information theoretic sense. Communication complexity studies such problems.

The general type of problem considered by communication complexity is as follows. Alice receives an input string $x \in \{0,1\}^n$ and Bob receives an input string $y \in \{0,1\}^n$. Their goal is to compute $f(x,y)$ for some function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. We can also consider the more general case where Alice and Bob receive strings of different lengths and the output of the function f is longer than just one bit, but the simpler model will be sufficient for this lecture. They wish to accomplish this task using as little communication as possible. We can measure the amount of communication in the number of bits or qubits sent, depending on whether we are considering classical or quantum communication complexity. For simplicity let us be more precise and say that Bob should learn the value $f(x,y)$—if instead we required that both Alice and Bob learn $f(x,y)$, the amount of communication required would differ by at most one bit. We are typically interested in the cost for the *worst case* choice of x and y, and will consider the minimum such cost over all possible protocols.

There are several variants of the communication complexity model that one may consider. We may consider quantum or classical communication complexity; we may require the answer to be correct with probability 1 or allow a small error probability ε that the wrong answer is found; and we may allow or disallow shared randomness or entanglement. Other variations, such as requiring the communication to go only one way or only allowing communication from Alice and Bob to a third party referee, are also frequently studied.

It is worth mentioning one formality of the communication complexity model, to make sure the model is properly understood. A protocol for a given function should completely specify the structure of the communication: Alice sends k_1 bits (or qubits) to Bob, Bob sends k_2 bits (or qubits) to Alice, Alice sends k_3 bits (or qubits) to Bob, and so on for some fixed number of rounds. The inputs x and y only influence the contents of these messages, not the sizes or timing of the messages or the number of rounds. This disallows complications such as "communicating by not communicating".

We can gain some insight about communication complexity by considering some example functions. The following three functions are important examples.

1. **Equality:**
$$\mathrm{EQ}(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y. \end{cases}$$

2. **Modulo 2 inner product:**
$$\mathrm{IP}(x,y) = x \cdot y.$$

3. **Disjointness:**
$$\mathrm{DISJ}(x,y) = \begin{cases} 1 & \text{if } x_i \wedge y_i = 1 \text{ for some } i \in \{1,\ldots,n\} \\ 0 & \text{otherwise.} \end{cases}$$

Let us first consider the classical communication complexity of the equality function. If we restrict our attention to deterministic protocols (that are correct with certainty), then nothing non-trivial is possible for equality: n bits of communication are both necessary and sufficient. We express this as
$$D(\mathrm{EQ}) = n.$$
In general, $D(f)$ denotes the deterministic communication complexity. Note that all functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ have $D(f) \leq n$, because Alice can always send her entire input x to Bob, who then can just evaluate $f(x,y)$. It is usually not too hard to put lower bounds on $D(f)$ for various functions f—I won't show you how to prove $D(\mathrm{EQ}) \geq n$, but you can probably convince yourself that it is true.

In the probabilistic case, a significant improvement is possible. A probabilistic protocol that has error bounded by $\varepsilon > 0$ exists that requires just $O(\log(n/\varepsilon))$ bits of communication. We will express this as
$$R(\mathrm{EQ}) = O(\log n)$$
(with $\varepsilon = 1/3$ being implicit if we do not specify the error). In fact, if we assume that Alice and Bob share $O(\log n)$ random bits, then only a constant number of bits of communication are required for constant error.

Let us briefly discuss how this is possible. The idea is to use an error correcting code with the right properties. Specifically, we want an error correcting code
$$E : \{0,1\}^n \to \{0,1\}^m$$
where $m = cn$ for some constant c, and
$$\mathrm{dist}(E(x), E(y)) \geq \delta m \quad \text{(for all } x \neq y)$$
for some constant $\delta > 0$. Such codes are known to exist.[1] Now, Alice and Bob both encode their strings using this code, Alice randomly chooses $i \in \{1,\ldots,m\}$, and sends $(E(x)_i, i)$ to Bob.

[1] I believe Justesen codes were the first examples of such codes. For these codes we can, for instance, take $c = 4$ and $\delta = 1/12$. I presume there are better constructions known.

This requires $O(\log m) = O(\log n)$ bits of communication. Bob compares $E(x)_i$ with $E(y)_i$. If $E(x)_i = E(y)_i$, Bob concludes $x = y$, and otherwise he concludes $x \neq y$. If $x = y$, Bob will be right, while if $x \neq y$ he will conclude with constant probability that $x \neq y$. The process can be repeated a constant number of times to decrease the probability of error. In the setting where Alice and Bob have shared randomness, the reduction in the cost to a constant number of bits comes from the fact that Alice does not need to send the index i to Bob.

For the other two example functions, it turns out that even using randomness and allowing a small probability of error, no asymptotic reduction in communication complexity is possible over the trivial protocol: we have $R(\mathrm{IP}) = \Theta(n)$ and $R(\mathrm{DISJ}) = \Theta(n)$. Unlike the typical situation for deterministic communication complexity, proving lower bounds for randomized protocols can be very difficult—so I will not attempt to prove to you that any protocol for disjointness has communication cost $\Omega(n)$ (but it is true).

Quantum communication complexity

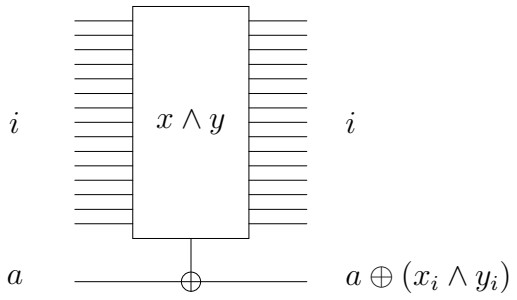
It is a natural question to ask whether quantum information can reduce communication complexity. In light of Holevo's Theorem, it may be tempting to think not, but in fact significant reductions are possible.

For example, let us show that the quantum communication complexity of disjointness, denoted $Q(\mathrm{DISJ})$, satisfies $Q(\mathrm{DISJ}) = O(\sqrt{n}\log n)$. More specifically, $Q(f)$ refers to quantum communication complexity allowing a small probability of error and disallowing shared entanglement. It turns out that $Q(\mathrm{DISJ}) = \Theta(\sqrt{n})$; the elimination of the logarithmic factor is clever optimization of the method we will see, while the lower bound requires a rather difficult proof.

Before describing the protocol, let us consider the following problem. Suppose that the strings x and y are fixed, but unknown to us, and imagine that we have access to a black box B_g for a function $g : \{0,1\}^n \to \{0,1\}$ satisfying

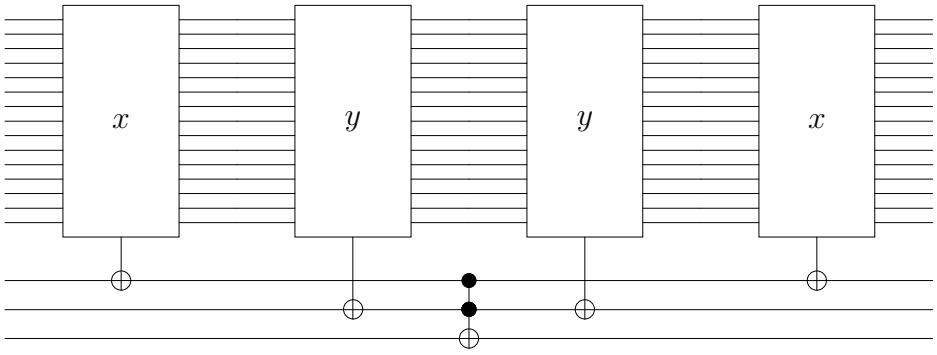
$$g(i) = x_i \wedge y_i.$$

As you would assume, the black box is defined on strings of length $k = \lceil \log_2 n \rceil$, each index i is encoded as a string of length k, and the black box implements the following reversible operation:



Using $O(\sqrt{n})$ queries to this black box, Grover's Algorithm could determine with high probability whether there exists an index i for which $g(i) = 1$ (which is equivalent to $x_i \wedge y_i = 1$).

Now, the way that Alice and Bob will solve the disjointness problem is essentially to run Grover's algorithm. Of course, neither of them has a black box for $x \wedge y$ as above, but they can implement this black box in a distributed fashion. Specifically, Alice can implement such a black box for x in place of $x \wedge y$ (because she knows x), and Bob can do likewise for y. Then they can combine their black boxes to simulate the one for $x \wedge y$ as follows:



Of course this requires communication: they must exchange $O(\log n)$ qubits in order to implement this circuit. However, that is good enough for what we need: Alice (say) can effectively run Grover's Algorithm to determine whether $x_i \wedge y_i = 1$ for some i, exchanging $O(\log n)$ qubits with Bob for each query. Because the total number of queries needed is at most $O(\sqrt{n})$, the total amount of communication is $O(\sqrt{n} \log n)$.

For other functions quantum information does not help. For instance, $Q(\mathrm{IP}) = \Theta(n)$. The idea behind the proof of this bound is that anything asymptotically less than $\Theta(n)$ would allow for a violation of Holevo's Theorem.

Exponential separation

It turns out that an exponential separation between quantum and classical communication complexity is possible, provided we allow for promises on the inputs. Ran Raz proved that the following problem gives such a separation.

Alice receives classical descriptions of (i) a unit vector $|\psi\rangle \in \mathbb{C}^n$ and (ii) a projective measurement $\{\Pi_0, \Pi_1\}$ on $\mathbb{C}^n$, and Bob receives a classical description of a unitary operator $U \in \mathrm{L}(\mathbb{C}^n)$. It is promised that one of $\|\Pi_0 U |\psi\rangle\|^2$ or $\|\Pi_1 U |\psi\rangle\|^2$ is close to 1 (and so the other is close to 0), and the goal is to output 0 or 1 accordingly. It is easy for Alice and Bob to solve this problem using $O(\log n)$ qubits of communication: Alice sends $|\psi\rangle$ to Bob, Bob applies U and sends it back to Alice, and Alice measures $U |\psi\rangle$ with respect to $\{\Pi_0, \Pi_1\}$. The challenging part is to prove that no classical probabilistic protocol can achieve this—it is possible to prove a lower bound of $\Omega(n^{1/4}/\log n)$ for the classical communication complexity. The proof is very difficult.

It is still an open question whether an exponential improvement of classical over quantum communication complexity is possible for a total function f (meaning that no promises are allowed in the corresponding problem).

Quantum fingerprinting

Finally, I would like to mention a variant of the communication complexity model where it is possible to prove an exponential separation between quantum and classical complexities for a total function (i.e., not a promise problem). It is a very weak model called the *simultaneous message passing model*. As before, Alice receives an input string x and Bob receives y, but now Alice and Bob cannot communicate with one another at all. Instead, they must both send a single message to a third party, called the referee. The referee does not receive any input, but is required to give the output of the function f based on the messages received from Alice and Bob.

One can consider several variants of this model—we will compare the classical probabilistic model with the quantum model, where the referee is allowed a small probability of error in both cases. No shared randomness or entanglement will be permitted between Alice and Bob. (If we do allow shared randomness or entanglement, the example to be considered no longer gives an interesting separation between quantum and classical.)

The separation will be for the equality function. If we write $R^{\|}(f)$ to denote the randomized communication complexity of f in the simultaneous message passing model, then we have

$$R^{\|}(\mathrm{EQ}) = \Theta(\sqrt{n}).$$

In contrast, using a similar notation for the quantum case, we have

$$Q^{\|}(\mathrm{EQ}) = O(\log n).$$

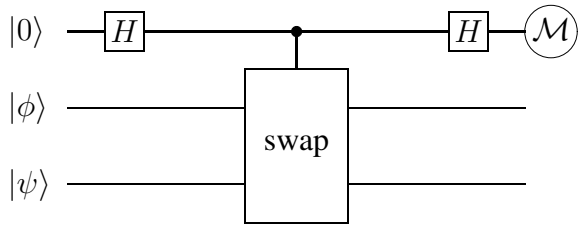
The technique that establishes that $Q^{\|}(\mathrm{EQ}) = O(\log n)$ is called *quantum fingerprinting*. The idea is that Alice will prepare a small quantum state $|\psi_x\rangle$ given x, and Bob will prepare $|\psi_y\rangle$ given y. We want that each $|\psi_x\rangle$ is a quantum state on a number of qubits that is logarithmic in n, but so that $|\psi_x\rangle$ and $|\psi_y\rangle$ are very different when $x \neq y$. The state $|\psi_x\rangle$ is a "fingerprint" of x—it does not contain enough information for someone to recover x from it, but it can be used to distinguish x from another string y with a different fingerprint. The referee will then attempt to check whether the fingerprints $|\psi_x\rangle$ and $|\psi_y\rangle$ are the same or different.

Of course it is not possible to come up with states $\{|\psi_x\rangle \,:\, x \in \{0,1\}^n\}$ that form an orthonormal set, because this would require that the fingerprints are too large: at least n qubits. So instead what we will do is to make them pairwise "almost orthogonal": $|\langle\psi_x|\psi_y\rangle|$ should be small whenever $x \neq y$. We can do this using exactly the same kind of error correcting codes we discussed in the first part of the lecture. Specifically, we let

$$|\psi_x\rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} |i\rangle \, |E(x)_i\rangle \,.$$

Then, assuming $\mathrm{dist}(E(x), E(y)) \geq \delta m$ for $x \neq y$, we have $|\langle\psi_x|\psi_y\rangle| \leq 1 - \delta$. Replacing $|\psi_x\rangle$ with some constant number of copies of $|\psi_x\rangle$ allows this bound on the inner product to be decreased to an arbitrarily small constant. The number of qubits required for these fingerprints is small: $O(\log n)$ qubits.

Now, how can the referee determine whether $|\psi_x\rangle = |\psi_y\rangle$ or $|\langle\psi_x|\psi_y\rangle|$ is small? You've already answered this question (in a slightly simplified form) in question 3 of homework assignment 1. In that question, you were asked to consider this circuit:



There were many ways to formulate you answer, but one way was to say

$$\Pr[\text{outcome is } 0] = \frac{1}{2} + \frac{|\langle\phi|\psi\rangle|^2}{2}$$

$$\Pr[\text{outcome is } 1] = \frac{1}{2} - \frac{|\langle\phi|\psi\rangle|^2}{2}.$$

This generalizes to states $|\phi\rangle$ and $|\psi\rangle$ on any number of qubits. Thus, the referee can simply perform this procedure (sometimes called the *swap test*) on $|\psi_x\rangle$ and $|\psi_y\rangle$. If he measures 0, he determines that $x = y$, and if he measures 1 he determines that $x \neq y$.

If it were the case that $x = y$, then the referee always answers correctly because he will always measure 0. If $x \neq y$, the referee might make a mistake and conclude incorrectly that $x = y$, but if Alice and Bob send some constant number of copies of their fingerprints, the probability of error can be reduced to any desired positive constant in the usual way.