Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Quantum Information Theory
Tips 11

FS 11
Lídia del Rio

Exercise 11.1 Entanglement and Teleportation

This exercise introduces a rather spectacular result of quantum information: if two parties, Alice and Bob, share an entangled state, than they can *teleport* a state from one side to the other at the cost of the entanglement between them.

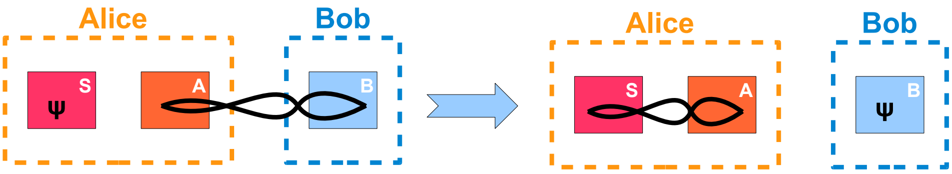

Figure 1: Quantum teleportation: in the beginning, Alice has a a qubit S in pure state $|\psi\rangle$ and a qubit A that is entangled with a qubit on Bob's side in a Bell state. By the end of the protocol, Alice's two qubits, S and A, will be entangled in a Bell state (not necessarily the same) and Bob's qubit, B, will be in state $|\psi\rangle$. The entanglement between Alice and Bob is broken when $|\psi\rangle$ is "teleported".

The setting is illustrated in Fig. 1. In her lab, Alice has a a qubit S in pure state $|\psi\rangle$ and a qubit A that is entangled with a qubit on Bob's side in a Bell state, $\frac{1}{\sqrt{2}}(|0_A 0_B\rangle + |1_A 1_B\rangle)$.
$|\psi\rangle$ is an arbitrary qubit pure state, so it may be written as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$. The global system, $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_B$, is initially in state

$$|\phi^0\rangle = (\alpha|0\rangle_S + \beta|1\rangle_S) \otimes \frac{1}{\sqrt{2}}(|0_A 0_B\rangle + |1_A 1_B\rangle) \,. \tag{1}$$

Now Alice measures her two qubits in the Bell basis,

$$\left\{|sa^k\rangle\right\}_k = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2}}(|0_S 0_A\rangle + |1_S 1_A\rangle)\,, & \frac{1}{\sqrt{2}}(|0_S 0_A\rangle - |1_S 1_A\rangle)\,, \\ \frac{1}{\sqrt{2}}(|0_S 1_A\rangle + |1_S 0_A\rangle)\,, & \frac{1}{\sqrt{2}}(|0_S 1_A\rangle - |1_S 0_A\rangle) \end{array} \right\}, \tag{2}$$

obtaining outcomes 1, 2, 3 and 4 for each of the states $|sa^k\rangle$ respectively. We will see that after her measurement Bob's qubit "collapses" to a sate *very close* to $|\psi\rangle$.
The first thing you should notice is that the projectors of that measurement include the identity on B, because she is not measuring anything on that system. For instance, the projector for the first state of the Bell basis, $|sa^1\rangle$, is

$$P_1 = \frac{1}{2}\big[\,(|0_S 0_A\rangle + |1_S 1_A\rangle)\,(\langle 0_S 0_A| + \langle 1_S 1_A|)\,\big] \otimes \mathbb{1}_B. \tag{3}$$

Let us see what happens when Alice measures that state on her qubits, i.e. obtains outcome 1. From the foundations of quantum mechanics (page 30 of the script) you know that the final state of the global system is

$$|\phi^1\rangle = \frac{P_1|\phi^0\rangle}{\sqrt{\Pr_1}}, \tag{4}$$

where $\Pr_1$ is the probability that the outcome of her measurement is 1. You can check that for this basis all outcomes are equally likely, $\Pr_k = \frac{1}{4}, \forall_k$. We obtain

$$\begin{aligned}
|\phi^1\rangle &= \frac{1}{\sqrt{2}}\Big[\left(|0_S 0_A\rangle + |1_S 1_A\rangle\right)\left(\langle 0_S 0_A| + \langle 1_S 1_A|\right)\otimes \mathbb{1}_B\Big]\Big[\left(\alpha|0\rangle_S + \beta|1\rangle_S\right)\otimes\left(|0_A 0_B\rangle + |1_A 1_B\rangle\right)\Big] \\
&= \frac{1}{\sqrt{2}}\left(|0_S 0_A\rangle + |1_S 1_A\rangle\right)\otimes\left[\begin{array}{c}
\langle \underline{0_S 0_A}| \otimes \mathbb{1}_B\Big[\Big(\underline{\alpha|0\rangle_S} + \beta|1\rangle_S\Big)\otimes\left(|\underline{0_A}0_B\rangle + |1_A 1_B\rangle\right)\Big] \\
\\
+\langle \underline{1_S 1_A}| \otimes \mathbb{1}_B\Big[\Big(\alpha|0\rangle_S + \underline{\beta|1\rangle_S}\Big)\otimes\left(|0_A 0_B\rangle + |\underline{1_A}1_B\rangle\right)\Big]
\end{array}\right] \\
&= \frac{1}{\sqrt{2}}\left(|0_S 0_A\rangle + |1_S 1_A\rangle\right)\otimes\big[\alpha|0\rangle_B + \beta|1\rangle_B\big] \quad = \quad |as^1\rangle \otimes |\psi\rangle_B.
\end{aligned}$$

I hope the rainbow above has not blinded you and that you managed to follow what happened there and how we ended up with a fully correlated Bell state on $S \otimes A$ that is decoupled from B, where we find $|\psi\rangle$. The key, of course, lies in the strong correlations between A and B. If you repeat this procedure to all possible outcomes(part a) of the exercise), you should obtain the table

Alice's outcome	Alice's state	Bob's state	Bob performs
1	$\lvert as^1\rangle = \frac{1}{\sqrt{2}}\left(\lvert 0_S 0_A\rangle + \lvert 1_S 1_A\rangle\right)$	$\lvert b^1\rangle = \alpha\lvert 0\rangle + \beta\lvert 1\rangle$	O_1
2	$\lvert as^2\rangle = \frac{1}{\sqrt{2}}\left(\lvert 0_S 0_A\rangle - \lvert 1_S 1_A\rangle\right)$	$\lvert b^2\rangle = \alpha\lvert 0\rangle - \beta\lvert 1\rangle$	$O_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
3	$\lvert as^3\rangle = \frac{1}{\sqrt{2}}\left(\lvert 0_S 1_A\rangle + \lvert 1_S 0_A\rangle\right)$	$\lvert b^3\rangle = \beta\lvert 0\rangle + \alpha\lvert 1\rangle$	O_3
4	$\lvert as^4\rangle = \frac{1}{\sqrt{2}}\left(\lvert 0_S 1_A\rangle - \lvert 1_S 0_A\rangle\right)$	$\lvert b^4\rangle = \beta\lvert 0\rangle - \alpha\lvert 1\rangle$	O_4

Not always it happens that the state of Bob's system is exactly $|\psi\rangle$. For instance, when Alice obtains outcome 2, his qubit goes to state $\alpha|0\rangle - \beta|1\rangle$, and he would have to perform a one-qubit operation on his system to recover $|\psi\rangle$. In this case, he would have to flip the sign of $|1\rangle$, applying the unitary represented by O_2 in the computational basis. For part b) you have to find all the other operations $\{O_k\}_k$.
Of course, Bob only knows what operations to apply because he knows the state $|b^k\rangle$ of his qubits, and he knows that because Alice told him the outcome k of her measurement. What if Alice had not told him the outcome? In that case, Bob would have to try to guess what the state of his qubit. He knows that all measurement outcomes were equally likely, and that for each of them he would have a different state. Fortunately, in quantum mechanics we have a way of describing probabilistic mixtures of pure states — with density matrices. The state Bob has after Alice's measurement is, from his point of view, $\rho = \sum_k \frac{1}{4}|b^k\rangle\langle b^k|$. In part c) you have to show that when Bob does not know the outcome of the measurement, he cannot have any idea of what his state is or how to recover $|\psi\rangle$, i.e. $\rho = \mathbb{1}_B$. This tells us that the quantum teleportation protocol can only work if Alice uses a (possibly classical) communication channel to share some information with Bob (the outcome of her measurement).
Notice that when Alice and Bob teleport the state of one qubit, they lose their entanglement, and therefore cannot repeat the protocol to teleport anything else. Impressive as it is, quantum teleportation comes with a cost. So far we have only seen how to teleport a pure state. One may wonder what happens if the state Alice tries to teleport is entangled with a reference system R that she does not control. Would the final state on Bob's side be entangled with R in the same way? The answer is, swimmingly, yes (Fig. 2).
In parts d) and e) of the exercise you are asked to prove that more formally. You can start by considering that every mixed state can be expanded in its eigenbasis, $\rho_S = \sum_i p_i |i\rangle\langle i|_S$, with $|i\rangle = \alpha_i|0\rangle + \beta_i|0\rangle$. Check that the protocol works for such a state. You can, for instance, show what happens when Alice measures her two qubits in the Bell basis and obtains outcome 2. Remember that the final state of the whole system is given by

$$\frac{1}{\Pr_2}\Big(|as^2\rangle\langle as^2| \otimes \mathbb{1}_B\Big)\left[\rho_S \otimes \frac{1}{\sqrt{2}}\left(|0_A 0_B\rangle + |1_A 1_B\rangle\right)\right]\Big(|as^2\rangle\langle as^2| \otimes \mathbb{1}_B\Big). \tag{5}$$

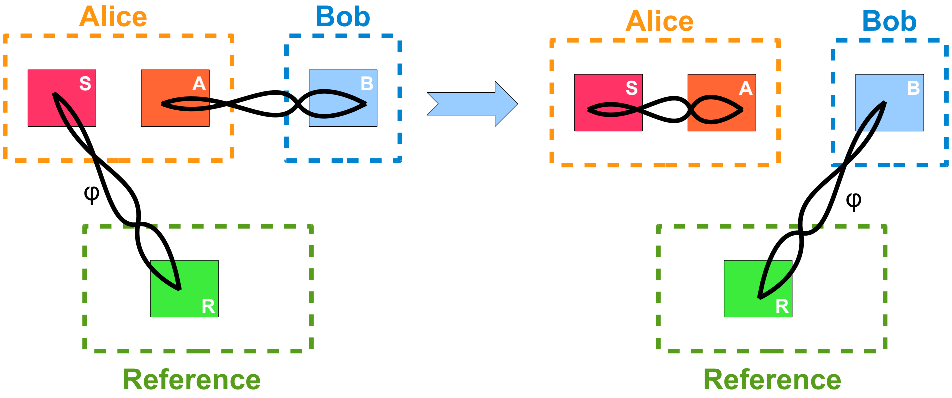

Figure 2: Quantum teleportation preserves entanglement. If Alice teleports a mixed state ρ_S that is entangled with a reference system R, $\rho_S = \text{Tr}_R|\phi\rangle\langle\phi|_{SR}$, not only the final state on Bob's side will be ρ but it will be entangled with R in the exact same way as before, $\rho_B = \text{Tr}_R|\phi\rangle\langle\phi|_{BR}$.

Verify that in order to recover ρ on his qubit, Bob only needs to apply the unitary O_2 as before. Argue that it also works for the remaining measurement outcomes. This implies, in particular, that the protocol is linear: you did not have to use the convexity of density operators ($\sum_i p_i = 1$) to prove this result. Now we are ready to tackle correlations between ρ_S and an external system R. Try making a Schmidt decomposition of the pure state $|\phi\rangle$ of $\mathcal{H}_S \otimes \mathcal{H}_R$. You should get something like $|\phi\rangle_{SR} = \sum_i p_i |i\rangle_S \otimes |i\rangle_R$. If you call the quantum teleportation protocol $\mathcal{E}$, apply $\mathcal{E} \otimes \mathcal{I}_R$ on that state and use the linearity of $\mathcal{E}$ you should obtain the result we are looking for.

Exercise 11.2 One-time pad

This is a purely classical exercise about encryption whose conclusions are important for quantum cryptography schemes. Imagine you were given a message and have to send it through a public, insecure channel. Because your message is super secret (for instance who you think should win Eurovision), you want to encrypt it as a ciphertext so that no one other than you, Alice, and your friend Bob can decipher it if they happen to be spying on your channel.
There are two main kinds of encryption schemes you may use:

- Symmetric: you and Bob share a key that is used to encode and decode the message. Examples are the one-time pad, that we will study here, and the (very insecure) shift cyphers.

- Asymmetric: you and Bob use different keys. Usually Bob shares a public key that you (or anyone else) can use to encrypt the message, but keeps a private key for himself and uses that one to decipher it. A well-known example is RSA, that relies on a computational hard problem to generate the keys — it is easy to obtain the public key if you have the private one but very hard to get the private key from the public key and thus decrypt the ciphertext. Most of the cryptography used currently in everyday life (eg. online shopping) is based on asymmetric schemes.

Suppose that in this case you want to use a symmetric scheme. We assume that because those are the schemes where quantum cryptography is more useful — to create, or rather expand, a shared key.
What requirements would you want your cryptographic protocol to fulfil? Well, ideally,

1. Someone who knows the encrypted message and the key, like Bob, can recover the original message perfectly.

2. An adversary who intercepts the encrypted message but does not know the key cannot get any information about the original message.

The first condition may be formalised using conditional entropies: you want the uncertainty about the message of someone who already knows the ciphertextand the secret key to be zero: $H(M|C, K) = 0$. As for

the second one, what we want is the encrypted and original messages to be completely decorrelated, i.e. to share no information: $I(M:C)=0$. If you work with these two propositions and use the fact that mutual information and entropy are always positive you should obtain the condition $H(K)\geq H(M)$.
What does this result mean? We know that entropy may be interpreted as randomness, so that condition would read "the randomness of the key has to be at least as large as the randomness of the message". But why? And how should we use the key to encrypt our message?
Maybe we should to illustrate this problem with a concreteexample to get some intuition to what is happening. Suppose that you want to encrypt two different messages,

$$M_1 = 001111001100110000, \qquad M_2 = 123789456234.$$

At first sight we detect a certain redundancy in both messages: in the first, all the bits are repeated; in the second there are blocks of three consecutive numbers. The first step of our encryption protocol could be simply to *compress* the original message so that as long as one knows how it was compressed recovering the original is trivial. This is not encrypting the message yet — we are just using the patterns of a message to decrease its size. In other words, we are eliminating all the redundancy possible and leaving only what seems random in the messages. In this case, we may use the lossless compression algorithms

$$\begin{aligned} &\mathcal{A}_1 : \{b,b\} \mapsto \{b\}\,; &&M_1' = 011010100,\\ &\mathcal{A}_2 : \{k,k+1,k+2\} \mapsto \{k\}\,; &&M_2' = 1742. \end{aligned}$$

Now the size of our messages is reduced to the essential and we can think of ways to encrypt it. Note a few things:

1. To be *lossless*, a compression algorithm has to be injective, so that we can always recover the original message from the compressed one.

2. The algorithms $\mathcal{A}_1$ and $\mathcal{A}_1$ can be made public as long as we encrypt M_1' as a ciphertextC_1 such that $I(M_1':C_1)=0$, and the same for M_2' — if one cannot guess the compressed message from the cyphertext, they could never guess the original message, even if they know the compression algorithms.

3. We can compress a message M until it has $H(M)$ bits. This comes from the definition of entropy as randomness — what is left when we eliminated all the patterns. Of course, different agents may see different patterns in the messages. Consider for instance the message

   $$M_3 = 111122335588,$$

   and suppose that two different agents want to compress it:

   - Agent A sees only that every number is repeated twice and compresses the message to 112358.
   - Agent B, who is over ten years old, notices that the message encodes a Fibonacci sequence where every number is repeated twice, and can compress it, for instance, to $1-1-6$ (where the first two numbers stand for the initial terms of the sequence and the last number, 6, for the total number of terms).

   Clearly the same message, M_3, was compressed to different sizes by the two agents. From the point of view of agent B, the original message had less randomness than agent B thought. This example should convince us that what we call the *objective* entropy of M, $H(M)$, is in fact the *subjective* entropy of the message *given the knowledge of the agent* that wants to compress it (for instance if they know the Fibonacci sequence), $H(M|A)$. Because there is usually only one agent, Alice, trying to compress the message, and we assume that she is infinitely clever and capable of spotting every hint of a pattern, we normally use only $H(M)$ instead of $H(M|A)$. However, you should keep in mind that this entropy is subjective and agent-dependent.

Now that Alice has a message M' of size $H(M)$, she needs to find a way to encrypt it. For simplicity, we assume that the compressed message is already written in bits. Consider $M'_1 = 011010100$. One example of a very simple encryption scheme would be to just add 1 $\pmod 2$ to every bit,

$$\begin{aligned} M' &= 011010100, \quad K = 1 \text{ (repeated)} \\ C &= 100101011, \end{aligned}$$

It is easy to understand that this scheme is very insecure - an adversary who intercepts the ciphertext$C$ only needs to test a one-bit key to obtain the message — they would know that they had used the right key because the final message they obtained after reverse engineering the key and compression algorithm should make same kind of sense, like translating to "come what may, do not vote for the UK!" or to a credit card number. More formally, the mutual information between the cypher and the message is non-zero, which means an adversary would not have to try very hard to recover the original message. If Alice had chosen a two-bit long key to encrypt it, the adversary would have needed to test every one-bit and two-bit keys before they could find a message that made sense — again, not very hard.

Now if Alice chooses a key *as long as the compressed message* at random and again adds every bit of it to the message (in $\mod 2$), something like

$$\begin{aligned} M' &= 011010100 \\ K &= 101011010 \\ C &= 110001110, \end{aligned}$$

an adversary that did not know the key used would have to test every string of the same size of the message to find the right key. This is in practice the same as testing all possible messages of the size of M', $H(M)$. It means that the only information the adversary has about the message is its size, and cannot obtain the right message by testing small keys.

In the exercise you prove formally, like Shannon did, what we have just shown in a intuitive way: that if the key is chosen totally at random (i.e. its entropy is as large as its length, or, in bits, $H(K) = \log |K|$), it needs to be at least as large as the compressed message, i.e., $H(K) \geq H(M)$.