Basics for Quantum Entanglement

Yi-Ming Ding, Westlake University

Updated on Nov 22, 2024

Contents

1 Introduction

This note summarizes some important and basic knowledge of quantum entanglement. The main references for the notes are

[1] M. A. Nielsen, I. L. Chuang. Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge: Cambridge University Press (2010).

[2] M. B. Plenio and S. Virmani. An introduction to entanglement measures. Quant. Inf. Comput.7:1-51 (2007).

[3] B. Zeng , X. Chen , DL Zhou , XG Wen. Quantum Information Meets Quantum Matter. Springer Press (2019).

2 Density Operator

2.1 Statistical view

Definition 1 *Suppose a quantum system is in one of a number of states* $\{\psi_j\}$ *with corresponding possibilities* $\{p_j\}$*, then the* ***density operator*** *of the system is defined as*

$$\hat{\rho} \equiv \sum_j p_j |\psi_j\rangle\langle\psi_j| \tag{1}$$

Statistically, we are considering an **ensemble** $\{p_j, |\psi_j\rangle\}$, which can be viewed as a generalization of the case where the system is in a certain state $|\psi_0\rangle$ with probability one. Notice that the probabilities here describe the uncertainty different from that in a superposition quantum state. The latter is actually a certain state as a whole.

Theorem 1 $\forall$ *density operator* $\hat{\rho}$ *and its corresponding Hilbert space* $\mathcal{H}$*,*

- ***Trace condition:*** $tr(\hat{\rho}) = 1$.
- ***Positivity condition:*** $\forall\, |\varphi\rangle \in \mathcal{H}$, $\langle\varphi|\hat{\rho}|\varphi\rangle \geq 0$.

2.2 Rephrasing quantum mechanics

Postulate 1 (State Space): *Associated to any isolated physical system is a complex vector space with inner product (that is,* ***a Hilbert space****) known as the* ***state space*** *of the system. The system is completely* *described by its **density operator**, which is a positive operator $\hat{\rho}$ with trace one, acting on the state space of the system. If a quantum system is in the state $\hat{\rho}_j$ with probability p_j, then the density operator for the system is $\sum_j p_j \hat{\rho}_j$.*

Postulate 2 (Evolution): *The **evolution** of a **closed** quantum system is described by a unitary transformation. That is, the state $\hat{\rho}$ of the system at time t_0 is related to the state $\hat{\rho}'$ of the system at time t_1 by a unitary operator $\hat{U}$ which depends only on the times t_0 and t_1,*

$$\hat{\rho}' = \hat{U}\hat{\rho}\hat{U}^\dagger \tag{2}$$

Postulate 3 (Quantum Measurement): ***Quantum measurements** are described by a collection $\{\hat{M}_m\}$ of **measurement operators**. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $\hat{\rho}$ immediately before the measurement then the probability that result m occurs is given by*

$$p(m) = \mathrm{tr}(\hat{M}^\dagger \hat{M} \hat{\rho}) \tag{3}$$

and the system after the measurement is

$$\frac{\hat{M}_m \hat{\rho} \hat{M}_m^\dagger}{\mathrm{tr}(\hat{M}_m^\dagger \hat{M}_m \hat{\rho})} \tag{4}$$

The measurement operators satisfy the completeness equation,

$$\sum_m \hat{M}_m^\dagger \hat{M}_m = \hat{I} \tag{5}$$

Postulate 4 (Composite System): *The state space of a **composite** physical system is the **tensor product** of the state spaces of the component physical systems. Moreover, if we have systems numbered 0 through $n-1$, and system number j is prepared in the state $\hat{\rho}_j$, then the joint state of the total system is*

$$\hat{\rho}_0 \otimes \hat{\rho}_1 \otimes \cdots \hat{\rho}_{n-1} \tag{6}$$

2.3 Pure state and mixed state

Definition 2 *A state is a **pure state** iff $\hat{\rho} = |\phi\rangle\langle\phi|$, i.e. the system is in a certain state with probability one. Otherwise, it is a **mixed state**.*

The key point here is that **can we** write $\hat{\rho}$ in the form of $|\phi\rangle\langle\phi|$ within some basis. For example, though

$$\hat{\rho} = \frac{1}{2}\Big(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|\Big) \tag{7}$$

looks like some mixed state, since

$$\hat{\rho} = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right) \left(\frac{\langle 0| + \langle 1|}{\sqrt{2}} \right) \tag{8}$$

this state is essentially a pure state.

Theorem 2 *$\forall$ density operator $\hat{\rho}$, we have $tr(\hat{\rho}^2) \leq 1$, with equality iff $\hat{\rho}$ is a pure state.*

2.4 Reduced density operator

Definition 3 *Suppose*

$$\hat{\rho}_{AB} = \sum_{jklm} p_{jklm} \Big(|j_A\rangle \otimes |k_B\rangle \Big) \Big(\langle l_A| \otimes \langle m_B| \Big) \tag{9}$$

*is the density operator of a **bipartite system** with two subsystem A and B, then the **reduced density operator** of A is defined as*

$$\hat{\rho}_A = tr_B(\hat{\rho}_{AB}) \equiv \sum_{jklm} p_{jklm} \Big(|j_A\rangle\langle l_A| \Big) tr \Big(|k_B\rangle\langle m_B| \Big) \tag{10}$$

*where the operation tr_B is called the **partial trace**.*

To better understand the concept of the reduced density operator, let us consider the simplest case, where A and B are two pure states, whose density operators are

$$\hat{\sigma}_A = |\phi_A\rangle\langle\phi_A|, \quad \hat{\sigma}_B = |\phi_B\rangle\langle\phi_B| \tag{11}$$

respectively. Then the density operator $\hat{\rho}_{AB}$ of the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ is

$$\hat{\rho}_{AB} = \hat{\sigma}_A \otimes \hat{\sigma}_B = \Big(|\phi_A\rangle \otimes |\phi_B\rangle \Big) \Big(\langle\phi_A| \otimes \langle\phi_B| \Big) \tag{12}$$

According to Def. 3, the reduced density operator of A should be

$$\hat{\rho}_A = |\phi_A\rangle\langle\phi_A| \mathrm{tr} \Big(|\phi_B\rangle\langle\phi_B| \Big) = |\phi_A\rangle\langle\phi_A| = \hat{\sigma}_A \tag{13}$$

which is exactly the density operator of A itself.

Now we go back to Def. 3. What the partial trace operation tr_B does in (10) is that it try removing the information of B in the bipartite system, thus leaves the information of A solely. Such a removal is based on tracing over the subsystem B only and cannot always be complete because $\hat{\rho}_A$ would include something remnant, the connection (entanglement) between A and B (See Fig. 1).

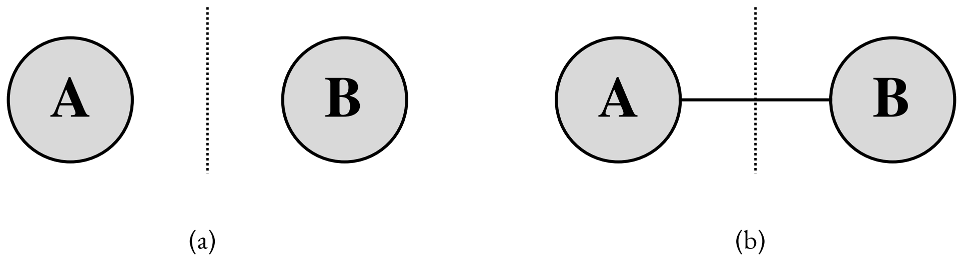

Figure 1: Let us consider a heuristic though not precise analogy: imagine A and B as two balls, and the partial trace operation tr_B as some knife that cuts the full system then throw B away. (a) If A and B are independent (no entanglements), $\hat{\rho}_A$ will be a complete description of A like Eq. (13); (b) if A and B are connected by a rope, the cut will leaves A and a half of the rope, which includes the information that how the two balls connected (entangled).

Let us take a look at another example. Suppose

$$|\psi_{AB}\rangle = \frac{|0_A 0_B\rangle + |1_A 1_B\rangle}{\sqrt{2}} \tag{14}$$

and

$$\begin{aligned}\hat{\rho}_{AB} =& |\psi_{AB}\rangle\langle\psi_{AB}| \\ =& \frac{|0_A 0_B\rangle\langle 0_A 0_B| + |0_A 0_B\rangle\langle 1_A 1_B| + |1_A 1_B\rangle\langle 0_A 0_B| + |1_A 1_B\rangle\langle 1_A 1_B|}{2}\end{aligned} \tag{15}$$

then

$$\hat{\rho}_A = \text{tr}_B \hat{\rho}_{AB} = \frac{|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|}{2} \tag{16}$$

It is easy to verify that $\hat{\rho}_A$ in Eq. (16) is a mixed state, which means that the state of $\hat{\rho}_A$ is in some uncertainty according to Def. 2. As we discussed above, since $\hat{\rho}_{AB}$ is a pure state, thus this uncertainty must come from the removal (trace) of B, which has a close relation with **quantum measurements**, which will be discussed in the next section.

2.5 Schmidt decomposition

Theorem 3 ***(Schmidt decomposition)*** *$\forall$ bipartite system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ and $|\Psi\rangle \in \mathcal{H}$ is a pure state, $\exists$ two sets of orthonormal basis $\{j_A\} \in \mathcal{H}_A$ and $\{j_B\} \in \mathcal{H}_B$, s.t.*

$$|\Psi\rangle = \sum_j \lambda_j |j_A\rangle |j_B\rangle \tag{17}$$

*where real numbers $\{\lambda_j\}$ are **Schmidt co-efficients**, which satisfy $\lambda_j \geq 0$ and $\sum_j \lambda_j^2 = 1$.*

Notice that $\hat{\rho}_A$ and $\hat{\rho}_B$ share the same $\{\lambda_j\}$, i.e. $\hat{\rho}_A = \sum_j \lambda_j^2 |j_A\rangle\langle j_A|$ and $\hat{\rho}_B = \sum_j \lambda_j^2 |j_B\rangle\langle j_B|$. Since the entanglement is symmetric to both sides, no matter on which side we do the partial trace, the terms that include their entanglement must be preserved and be the same. This is exactly the physical meaning of $\{\lambda_j\}$. Another intriguing thing is that for $\hat{\rho}_A$, Schmidt decomposition provides $\hat{\rho}_A$'s spectrum decomposition $\sum_j \lambda_j^2 |j_A\rangle\langle j_A|$ with the spectrum $\{\lambda_j^2\}$.

2.6 Relation between partial trace and measurement

Consider a Schmidt decomposed state (17), after we measure B under the basis $\{j_B\}$, the system will collapse to $|j_A\rangle \otimes |j_B\rangle$ with probability λ_j^2. If we have no idea what the outcome of the measurement gives, say we discard B after the meausrement without reading the results, then the state of A must be described statistically, i.e.

$$\hat{\rho}_A = \sum_j \lambda_j^2 |j_A\rangle\langle j_A| \tag{18}$$

On another hand, if we compute the reduced density operator of A, we will get

$$\begin{aligned} \mathrm{tr}_B(\hat{\rho}_{AB}) =& \mathrm{tr}_B\left(\sum_{jk} \lambda_j \lambda_k |j_A\rangle\langle k_A| \otimes |j_B\rangle\langle k_B|\right) \\ =& \sum_j \lambda_j^2 |j_A\rangle\langle j_A| \end{aligned} \tag{19}$$

which is exactly the operation of *measure and discard*.

Just as Def. 1, the reduced density operator is also from a statistical concept. Then in which case $\hat{\rho}_A$ is a mixed state? Obviously, if $\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B$, i.e. $\lambda_0 = 1$ and $\lambda_j = 0$ $(j \neq 0)$ in Eq. (17), the measurements on B will not affect A. However, if it is not the case, when A and B are entangled, the measurements on B will determine the state of A. This is exactly where the uncertainty of A comes from — the entanglement between A and B.

3 Entropies for entanglement

3.1 Shannon entropy

How to describe the amount of information? The genius insight proposed by C. Shannon is that *less uncertainty an event X has, more information we have for it before we measure the outcome*. For example, let X be coin tossing. If the two sides of the coin are exactly the same, such as a puppy pattern, then we will get the puppy pattern with probability 100%. In this case, the uncertainty is 0, and we know everything (maximum information) about the outcome of X even we do not do the toss. Similarly, if one of the sides is replaced by kitty pattern, then the outcome of X is uncertain, in which case we know less about the outcome unless we toss it because it will be the puppy pattern with some probability p and the kitty pattern with some probability $(1-p)$.

Then how to quantify the total information an event has? Suppose an event X, a random variable, can take d different independent values $\{x_j \mid j = 0, \cdots d-1\}$ with probability $\{p_j \mid j = 0, \cdots d-1\}$ to occur respectively. For each x_j, suppose its contribution to the total information is $I(p_j)$, which shall only depends on its probability, then the total information H is the average of different $I(p_j)$, i.e.

$$H(X) = \sum_j p_j I(p_j) \tag{20}$$

We require $I(p_j)$ satisfy the following conditions:

- $I(p)$ is a smooth function of probability with $p \in [0, 1]$.
- $I(pq) = I(p) + I(q)$.

Therefore

$$I(p) = k \log p \tag{21}$$

For convenience, we require $k = -1$ to make $I(p)$ non-negative. Therefore

$$H(X) = -\sum_j p_j \log p_j \tag{22}$$

which is called the **Shannon entropy**. Under this definition, $0 \log 0 \equiv 0$ is also reasonable. Note that if $p_0 = 1$ and $p_j = 0$ $(j \neq 0)$, then $H(X) = 0$, which means $H(X)$ is basically something describe the uncertainty. In other words, *the smaller $H(X)$, more information we have for X before we measure the outcome*.

Theorem 4 $H(X) \leq \log d$. *The equality holds* iff $p_0 = p_1 = \cdots p_{d-1} = 1/d$.

3.2 Min-entropy and Rényi entropy

Definition 4 *Suppose random variable X has d different outcomes $\{x_j \mid j = 0, 1, \cdots d-1\}$ and the corresponding probabilities are $\{p_j \mid j = 0, 1 \cdots d-1\}$. Define the **min-entropy** as*

$$H_{\min}(X) = -\log p_{\max}, \quad p_{\max} \equiv \max_j p_j \tag{23}$$

The difference between Shannon entropy and min-entropy is that Shannon entropy take all possibilities into accounts as an average while the min-entropy is much more conservative, which only measure the information of most possible outcome. On another hand, we know nothing about the information of $p_{\max}$ from the Shannon entropy because of the average. From this view, min-entropy and Shannon entropy are two extremes analogous to the ∞-norm and 1-norm. To achieve a more general measure of uncertainty and a tradeoff between the min-entropy and the Shannon entropy just like the p-norm, we give the following definition.

Definition 5 *Suppose random variable X has d different outcomes $\{x_j \mid j = 0, 1, \cdots d-1\}$ and the corresponding probabilities are $\{p_j \mid j = 0, 1 \cdots d-1\}$. Define the* ***Rényi entripy*** *as*

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_j p_j^\alpha \right) \tag{24}$$

Theorem 5 *Rényi entropy as a generalization of different entropies:*

- *If $\alpha = 1$, $H_\alpha(X) = H(X)$;*
- *If $\alpha = \infty$, $H_\alpha(X) = H_{\min}(X)$;*

3.3 Mutual information

Consider two random variables $X \in \{x_j\}$ and $Y \in \{y_j\}$. Suppose the probability of $X = x_j$ is $p(x_j)$ and the probability of $Y = y_k$ is $p(y_k)$. The joint probability of $(X, Y) = (x_j, y_k)$ is $p(x_j, y_k)$.

Definition 6 ***(Joint entropy)***

$$H(X, Y) \equiv -\sum_{x_j, y_k} p(x_j, y_k) \log p(x_j, y_k) \tag{25}$$

Definition 7 ***(Conditional entropy)***

$$H(X|Y) \equiv H(X, Y) - H(Y) \tag{26}$$

where $p(x_j|y_k)$ denotes the conditional probability.

Definition 8 ***(Mutual information)***

$$\begin{aligned} I(X:Y) \equiv& H(X) + H(Y) - H(X, Y) \\ =& H(X) - H(X|Y) = H(Y) - H(Y|X) \end{aligned} \tag{27}$$

The mutual information describes the uncertainty determined by X and Y together, or the information that they share in common. Different from the correlation function which usually depends on the knowledge of the related observables, the mutual information provides a more general way to quantify the correlation between two subsystems.

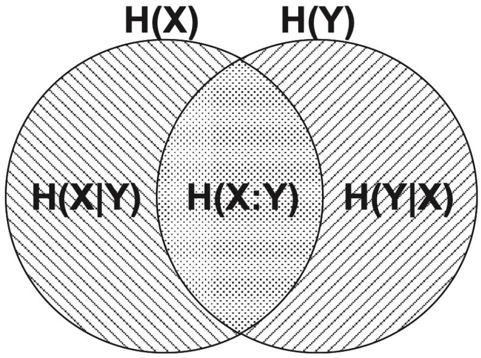

Figure 2: Relations among entropy $H(X)$, $H(Y)$, joint entropy $H(X,Y)$, conditional entropy $H(X|Y)$, $H(Y|X)$ and mutual information $H(X:Y)$. Picture from [*QCQI, M.A. Nielsen & I.L. Chuang*].

3.4 Quantum entropy

Definition 9 *Given a quantum state $\hat{\rho}$, the* ***Von Neumann entropy*** *$S(\hat{\rho})$ of it is defined as*

$$S(\hat{\rho}) \equiv -tr(\hat{\rho}\log\hat{\rho}) \tag{28}$$

Since a mixed state includes uncertainty, it is a natural idea to use entropy to describe the uncertainty (information) of a quantum state $\hat{\rho}$. Given a n-qubit quantum state

$$\hat{\rho} = \sum_j p_j |\psi_j\rangle\langle\psi_j| \tag{29}$$

where $\{|\psi_j\rangle\}$ is a orthogonal basis, then its matrix representation is a diagonal matrix diag$\{p_0, p_1, \cdots, p_{d-1}\}$ with $d = 2^n$. In this case, we similarly calculate the entropy

$$-\sum_j p_j \log p_j \equiv -\mathrm{tr}(\hat{\rho}\log\hat{\rho}) \tag{30}$$

which is exactly our definition of Von Neumann entropy. Therefore, the Von Neumann entropy is just the quantum generalization of the Shannon entropy. The trace operation guarantees the Von Neumann entropy shall be independent of the choice of representation, which is reasonable.

Definition 10 *A quantum state $\hat{\rho}$ is a* ***maximally entangled state*** *iff $S(\hat{\rho}) = \log d$, where d is the dimension of its Hilbert space.*

Apparently, the Bell states are maximally entangled state.

Definition 11 *Given a quantum state $\hat{\rho}$, the* ***(quantum) Rényi entropy*** *$S_\alpha(\hat{\rho})$ of it is defined as*

$$S_\alpha(\hat{\rho}) \equiv \frac{1}{1-\alpha}\log\left[tr(\rho^\alpha)\right] \tag{31}$$

Theorem 6 *Rényi entropy as a generalization of Von Neumann entropy:*

- *If* $\alpha = 1$, $S_\alpha(\hat{\rho}) = S(\hat{\rho})$;

3.5 Entanglement entropy

In a bipartite pure system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, the reduced density operator $\hat{\rho}_A$ and $\hat{\rho}_B$ of subsystems A and B respectively are some pure states *iff* there is no entanglements between them. From the definition of entropy, $S_\alpha(\hat{\rho}) = 0$ *iff* $\hat{\rho}$ is a pure state. Therefore, the quantum entropy can be viewed as a measure of entanglement. In this context, we give it the name the **entanglement entropy** (Von Neumann entropy or other Rényi entropy).

3.6 Entanglement from the view of mutual information

Definition 12 *Consider a pure state in the bipartite system* $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. *The* ***mutual information*** *between the subsystem A and B is defined as*

$$S(A : B) \equiv S(A) + S(B) - S(A, B) \tag{32}$$

In the classical case, if A and B are perfect correlated, then from Eq. (27) and Fig 2, $H(A) = H(B) = H(A, B)$, i.e.

$$H(A : B) = H(A) = H(B) = H(A, B) \tag{33}$$

However, in the quantum case,

$$S(A, B) = 2S(A) = 2S(B) \tag{34}$$

since $S(A, B) = 0$, which is twice of its classical counterpart.

Where is the factor 2 comes from? One may first expect there are some hidden-variables excluded, which contribute to the extra uncertainty. This kind of opinion corresponds to the **hidden-variable theory**, has been experimentally disprove with the well-known **Bell inequality**. In other words, quantum systems must include some quantum correlation, dubbed the **entanglement**, which the classical systems do not possess.