maximally entangled states: we will learn soon that *the degree of entanglement can be quantified by how mixed the reduced state is.*

Now let us ask what is the state of A after we measure B. As we have seen in Eq. (2.57), the composite state after the measurement can be either $|0,0\rangle$ or $|1,1\rangle$, both occurring with probability 1/2. Thus, if Alice does not know the outcomes of the measurements that B performed, then best possible guess to the state of A will be a classical probabilistic combination

$$\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.64}$$

which is exactly the same state as (2.63). Hence, from the point of view of Alice, it is impossible to know if the state of A is mixed because of entanglement or if it is mixed because Bob performed some measurements. All Alice can know is that the density matrix of A has the form of a maximally mixed state. This is called **the ambiguity of mixtures**. Even though the global AB state is affected by the measurement, from the point of view of Alice, she has no way of knowing. The only way that A would know is if she receives a *classical communication* from B. That is, if Bob sends an e-mail to Alice saying "Hey Alice, are you going to the party tonight? Oh, by the way, I measured my qubit and found it in 0."

2.6 The partial trace

The calculation that led us to Eq. (2.61) is what we call a partial trace. The trace, which we studied in Sec. 1.11, is an operation that receives an operator and spits out a number. The partial trace is an operation which receives a tensor product of operators and spits another operator, but living in a smaller Hilbert space. Why this is the correct procedure to be used in defining a reduced density matrix will be explained shortly.

Consider again a composite system AB. Let $|a\rangle$ and $|b\rangle$ be basis sets for A and B. Then a possible basis for AB is the tensor basis $|a,b\rangle$. What I want to do is investigate the trace operation within the full AB space. To do that, let us consider a general operator of the form $\mathcal{O} = A \otimes B$. After we learn how to deal with this, then we can generalize for an arbitrary operator, since any operator on AB can always be written as

$$\mathcal{O} = \sum_\alpha A_\alpha \otimes B_\alpha \tag{2.65}$$

for some index α and some set of operators A_α and B_α.

Let us then compute the trace of $\mathcal{O} = A \otimes B$ in the $|a,b\rangle$ basis:

$$\begin{aligned}
\operatorname{tr}(\mathcal{O}) &= \sum_{a,b} \langle a,b|\mathcal{O}|a,b\rangle \\
&= \sum_{a,b} (\langle a| \otimes \langle b|)(A \otimes B)(|a\rangle \otimes |b\rangle) \\
&= \sum_{a,b} \langle a|A|a\rangle \otimes \langle b|B|b\rangle \\
&= \sum_{a} \langle a|A|a\rangle \sum_{b} \langle b|B|b\rangle
\end{aligned}$$

I got rid of the $\otimes$ in the last line because the kron of two numbers is a number. The two terms in this formula are simply the trace of the operators A and B in their respective Hilbert spaces. Whence, we conclude that

$$\boxed{\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B)} \tag{2.66}$$

Now we can imagine an operation where we only trace over a part of the system. This is what we call the **partial trace**. It is defined as

$$\boxed{\operatorname{tr}_A(A \otimes B) = \operatorname{tr}(A) B, \qquad \operatorname{tr}_B(A \otimes B) = A \operatorname{tr}(B)} \tag{2.67}$$

When you "trace over A", you eliminate the variables pertaining to A and what you get left is an operator acting only on $\mathscr{H}_B$. This is something we often forget, so please pay attention: the result of a partial trace is still an operator. More generally, for an arbitrary operator $\mathcal{O}$ as defined in Eq. (2.65), we have

$$\operatorname{tr}_A \mathcal{O} = \sum_\alpha \operatorname{tr}(A_\alpha) B_\alpha \qquad \operatorname{tr}_B \mathcal{O} = \sum_\alpha A_\alpha \operatorname{tr}(B_\alpha) \tag{2.68}$$

As an example, suppose we have two qubits, with Pauli operators σ_x^A and σ_x^B. Then we would have, for instance,

$$\operatorname{tr}_A(\sigma_x^A \sigma_x^B) = \operatorname{tr}(\sigma_x)\sigma_x^B$$

Note how in the right-hand side I wrote σ_x instead of σ_x^A. The partial trace acts only on the single-spin subspace, so it does not matter which notation I use. Of course, this example I just gave is a bit silly because $\operatorname{tr}(\sigma_x) = 0$. But still, you get the idea. As another example, consider the partial trace of $\boldsymbol{\sigma}_A \cdot \boldsymbol{\sigma}_B = \sigma_x^A \sigma_x^B + \sigma_y^A \sigma_y^B + \sigma_z^A \sigma_z^B$. To compute it we need to use the linearity of the trace:

$$\operatorname{tr}_A(\boldsymbol{\sigma}_A \cdot \boldsymbol{\sigma}_B) = \operatorname{tr}(\sigma_x)\sigma_B^x + \operatorname{tr}(\sigma_y)\sigma_B^y + \operatorname{tr}(\sigma_z)\sigma_B^z$$

Again, all terms are zero in the end, so sorry again for the silly example. In principle every operator may be written in the form (2.65) so linearity solves all problems. However, that does not mean that writing down such an expansion is easy. For instance, suppose you want to compute the partial trace of $e^{\boldsymbol{\sigma}_A\cdot\boldsymbol{\sigma}_B}$. This turns out to be a quite clumsy calculation. For two qubits the matrices will be 4×4, so albeit clumsy, this is something a computer can readily do. For N qubits things become more difficult.

We can also write down the partial trace in terms of components. For instance, the partial trace over B reads:

$$\boxed{\operatorname{tr}_B \mathcal{O} = \sum_b \langle b|\mathcal{O}|b\rangle} \tag{2.69}$$

This notation may be a bit confusing at first. Actually, when we write $|b\rangle$ here, what we really mean is $1\otimes|b\rangle$. So the full formula would be

$$\operatorname{tr}_B \mathcal{O} = \sum_b (1\otimes\langle b|)\mathcal{O}(1\otimes|b\rangle) \tag{2.70}$$

We can check that this works using $\mathcal{O} = A\otimes B$. We then get

$$\begin{aligned}\operatorname{tr}_B \mathcal{O} &= \sum_b (1\otimes\langle b|)(A\otimes B)(1\otimes|b\rangle)\\ &= \sum_b (1A1)\otimes(\langle b|B|b\rangle)\\ &= A\sum_b \langle b|B|b\rangle\\ &= A\operatorname{tr}(B)\end{aligned}$$

Eq. (2.69) with $1\otimes|b\rangle$ is a convenient way to implement the partial trace in a computer.

Finally we could also write down a general formula for the partial trace in terms of the components of $\mathcal{O}$ in a basis. To do that, note that we may always insert two identities to decompose $\mathcal{O}$ as

$$\mathcal{O} = \sum_{a,b,a',b'} |a,b\rangle\langle a,b|\mathcal{O}|a',b'\rangle\langle a',b'| \tag{2.71}$$

To perform the partial trace over B, for instance, we sum over the diagonal entries of the B part $(b'=b)$:

$$\operatorname{tr}_B \mathcal{O} = \sum_{a,b,a'} |a\rangle\langle a,b|\mathcal{O}|a',b\rangle\langle a'| \tag{2.72}$$

The result is an operator acting on A, which we can see from the fact that this is a sum of outer products of the form $|a\rangle\langle a'|$:

$$\boxed{\operatorname{tr}_B \mathcal{O} = \sum_{a,a'} \left[\sum_b \langle a, b|\mathcal{O}|a', b\rangle\right] |a\rangle\langle a'|} \tag{2.73}$$

An example that is often encountered is the partial trace of some outer product, such as $|a, b\rangle\langle a', b'|$. To take the partial trace, remember that this can be written as

$$|a, b\rangle\langle a', b'| = |a\rangle\langle a'| \otimes |b\rangle\langle b'|$$

The partial trace over B, for instance, will simply go right through the first part and act only on the second part; i.e.,

$$\begin{aligned}\operatorname{tr}_B |a, b\rangle\langle a', b'| &= |a\rangle\langle a'| \operatorname{tr} \left\{|b\rangle\langle b'|\right\} \\ &= |a\rangle\langle a'|\left\{\langle b'|b\rangle\right\}\end{aligned}$$

Thus, we conclude that

$$\boxed{\operatorname{tr}_A |a, b\rangle\langle a', b'| = \delta_{a,a'} |b\rangle\langle b'|, \qquad \operatorname{tr}_B |a, b\rangle\langle a', b'| = |a\rangle\langle a'|\delta_{b,b'}} \tag{2.74}$$

2.7 Reduced density matrices

We are now ready to introduce the idea of a reduced density matrix in a more formal way. Given a bipartite system ρ_{AB} we define the reduced density matrix of A and B as

$$\boxed{\rho_A = \operatorname{tr}_B \rho_{AB}, \qquad \rho_B = \operatorname{tr}_A \rho_{AB}} \tag{2.75}$$

Thus, with the tools described in the previous section, it is now a matter of practice to play around and find reduced density matrices. It is also important to note that the partial trace works for both pure and mixed states. If we are dealing with pure states, then we simply write the density matrix as $|\psi\rangle\langle\psi|$ and continue as usual.

To warm up consider again the Bell state example that led us from the bipartite state (2.48) to the reduced state (2.63). Then

$$\rho_{AB} = |\Phi_1\rangle\langle\Phi_1| = \frac{1}{2}\left\{|0, 0\rangle\langle 0, 0| + |0, 0\rangle\langle 1, 1| + |1, 1\rangle\langle 0, 0| + |1, 1\rangle\langle 1, 1|\right\} \tag{2.76}$$

To take the partial trace we use Eq. (2.74). We then get

$$\rho_A = \frac{1}{2}\bigg\{|0\rangle\langle 0| + |1\rangle\langle 1|\bigg\} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.77}$$

with an identical result for ρ_B.

Let us look at some further examples. For instance, if we have a state which is of the form $\rho_{AB} = \rho_A \otimes \rho_B$, then Eq. (2.67) directly gives us $\operatorname{tr}_B \rho_{AB} = \rho_A$ and $\operatorname{tr}_A \rho_{AB} = \rho_B$, as of course expected. So any density matrix which is a product of the form $\rho_{AB} = \rho_A \otimes \rho_B$ represents *uncorrelated* systems, irrespective of whether the state is pure or not. However, it is very important to note that in general we *cannot* recover the full density matrix ρ_{AB} from the reduced density matrices ρ_A and ρ_B. The operation of taking the partial trace is **irreversible** and in general looses information. To put that more precisely, given a general ρ_{AB} and its reduced density matrices (2.75), we have

$$\rho_A \otimes \rho_B \neq \rho_{AB} \tag{2.78}$$

This is only true when ρ_{AB} was already originally uncorrelated. Thus, in general, we see that information is lost whenever AB are correlated.

To given an example, suppose we have two qubits in a state of the form

$$\rho_{AB} = \rho_A^0 \otimes \rho_B^0 + \chi \tag{2.79}$$

where

$$\chi = \alpha\bigg\{|0,1\rangle\langle 1,0| + |1,0\rangle\langle 0,1|\bigg\} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.80}$$

with α being a parameter.[2] What I like about (2.79) is that the partial trace of χ is always zero: $\operatorname{tr}_A(\chi) = \operatorname{tr}_B(\chi) = 0$. Thus, the reduced density matrices are $\rho_A = \rho_A^0$ and $\rho_B = \rho_B^0$. This means that from the perspective of A and B, it is as if χ doesn't even exist. But from a *global* perspective, you have a certain degree of correlations.

The partial trace is the quantum analog of *marginalizing* a probability distribution. To see that in first hand, consider a bipartite state of the form

$$\rho_{AB} = \sum_{i,j} p_{i,j} |i,j\rangle\langle i,j| \tag{2.81}$$

which will be a valid quantum state provided $p_{i,j} \in [0,1]$ and $\sum_{i,j} p_{i,j} = 1$. This state is as close as one gets from a *classical* probability distribution. To compute the partial trace over B we use Eq. (2.74), which gives

$$\rho_A = \operatorname{tr}_B \rho_{AB} = \sum_{i,j} p_{i,j} |i\rangle\langle i| = \sum_i p_i |i\rangle\langle i|$$

---

[2] The allowed values of α depend on ρ_A^0 and ρ_B^0 in order for te purity to be within the bounds $\frac{1}{4} \leq \mathcal{P} \leq 1$.

In the last equality I carried out the sum over j and defined

$$p_i = \sum_j p_{i,j} \tag{2.82}$$

This is exactly the marginalization procedure in classical probability theory. We simply sum over all probabilities of B to obtain a reduced probability distribution only for A.

Finally, let us talk about *why* this partial trace operation works. Or, putting it more precisely, why does the rule (2.67) works. What we really require of the partial trace operation is that

$$\mathrm{tr}_{\mathrm{AB}}\bigg[(A\otimes 1)\rho_{AB}\bigg] = \mathrm{tr}_A\bigg[A\rho_A\bigg] \tag{2.83}$$

That is, taking expectation values of A operators over the full Hilbert space or over the reduced Hilbert space give the same result. This is clearly true for the partial trace as we defined. What is a bit more subtle, is to show that the partial trace is the unique operation satisfying this criteria. This is demonstrated in Sec. 2.4.3 of Nielsen and Chuang.

2.8 Singular value and Schmidt decompositions

Consider a bipartite system AB with basis $|a,b\rangle$. The most general pure state in this system can be written as

$$|\psi\rangle = \sum_{a,b}\psi_{ab}|a,b\rangle, \tag{2.84}$$

where ψ_{ab} are coefficients. This state will in general be entangled. To see that in first hand, let us look at the reduced density matrices of A and B. I will leave for you as an exercise to show that

$$\rho_A = \mathrm{tr}_B\,|\psi\rangle\langle\psi| = \sum_{a,a'}\bigg[\sum_b \psi^*_{ab}\psi_{a'b}\bigg]|a\rangle\langle a'|, \tag{2.85}$$

$$\rho_B = \mathrm{tr}_A\,|\psi\rangle\langle\psi| = \sum_{b,b'}\bigg[\sum_a \psi^*_{ab}\psi_{ab'}\bigg]|b\rangle\langle b'|. \tag{2.86}$$

Of course, these are kind of ugly because ρ_A and ρ_B are not diagonal. But what I want to stress is that in general these states will be mixed. The only case in which these states will be pure is when the $\psi_{a,b}$ factor as a product of coefficients $\psi_{a,b} = f_a g_b$. Then one can already see from (2.84) that $|\psi\rangle$ will also factor as a product.

Suppose we now have a N-partite system with basis elements $|s_1,\dots,s_N\rangle$. Then the most general state pure state of this system will be

$$|\psi\rangle = \sum_{s_1,\dots,s_N}\psi_{s_1\dots s_N}|s_1,\dots,s_N\rangle. \tag{2.87}$$

The coefficients $\psi_{s_1 \ldots s_N}$ contain all the information about this system. It says, for instance, that 3 is entangled to 25 but 1 is not entangled with 12. Or that $1, 2, 3$ taken as a block, is completely independent of $4, 5, \ldots, N$. Everything is encoded in $\psi_{s_1, \ldots, s_N}$. Understanding how to extract the physics from this messy ψ is one of the most important questions of modern research.

If we think about it for a second, we also see that $\psi_{s_1, \ldots, s_N}$ can be viewed as a **tensor**. It is a rank-N tensor where each index has dimension d (the dimension of the local Hilbert space). Thus, there are in total d^N possible entries in this tensor. The physics of the state is then encoded inside this very very messy tensor structure. And that is a big problem because for $d = 2$ and $N = 300$, 2^{300} represents more particles than there are in the universe. Thus, if we want to characterize the entanglement properties of only 300 qubits, we are already in huge trouble because this is not a computational limitation that will be solved with the next generation of processors. It is a fundamental constraint.

The difficulties underlying the complex entanglement structure of states such as (2.87) has given rise to a new field of research known as **tensor networks**. The idea is two-fold. First, to create tools (such as diagrams and numerical libraries) which are efficient at dealing with complex tensors and give us intuition on what to do. Second, and most importantly, to understand what types of tensor structures appear most often. You see, *the many-body Hilbert space is enormous, but that does not mean that all of it is equally important.* It may very well be that in most typical scenarios, only a small part of the full Hilbert space is occupied. Figuring out what parts of the many-body Hilbert space are relevant is a million dollar question. Substantial progress has been done recently for certain classes of quantum systems, such as one-dimensional chains with short-range interactions. But the problem is nonetheless still in its infancy.

Singular Value Decomposition

In this section we will introduce some tools for dealing with the entanglement properties of quantum states. We start with a linear algebra tool that also has applications in many other fields of research, called the **singular value decomposition (SVD)**. Twenty years ago no one would teach the SVD for undergraduates. In twenty years from now, I guarantee you, SVD will be standard textbook material. The SVD theorem is as follows. Let A be an arbitrary rectangular $M \times N$ matrix. Then it is always possible to decompose A as

$$\boxed{A = USV^\dagger,} \tag{2.88}$$

where

- U is $M \times \min(M, N)$ and has *orthogonal columns* $U^\dagger U = 1$. If $M \leq N$ then U will be square and unitary, $UU^\dagger = 1$.
- V is $N \times \min(M, N)$ and has *orthogonal columns* $V^\dagger V = 1$. If $M \geq N$ then V will be square and unitary, $VV^\dagger = 1$.

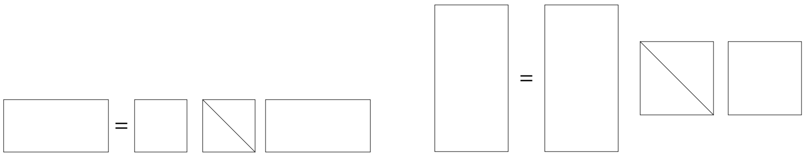

Figure 2.2: The size of the matrices appearing in Eq. (2.88). Left: A is short and fat ($M \leq N$). Right: A is thin and tall ($M \geq N$).

- S is $\min(M, N) \times \min(M, N)$ and diagonal, with entries $S_{\alpha\alpha} = \sigma_\alpha \geq 0$, which are called the **singular values** of the matrix A. It is convention to always order the singular values in decreasing order, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$. The number of non-zero singular values, called r, is known as the *Schmidt rank* of the matrix.

When the matrix is square, $M = N$, then both U and V become unitary. The sizes of A, U, S and V are shown in Fig. 2.2. For future reference, I will also write down Eq. (2.88) in terms of the components of A:

$$A_{ij} = \sum_{\alpha=1}^{r} U_{i\alpha} \sigma_\alpha V^*_{j\alpha} \tag{2.89}$$

where the sum extends only up the Schmidt rank r (after that the σ_α are zero so we don't need to include them).

The SVD is not in general related to eigenvalues of A. In fact, it is defined even for rectangular matrices. Instead, the SVD is actually related to the eigenvalues of $A^\dagger A$ and $AA^\dagger$. Starting from Eq. (2.88) and using the fact that $U^\dagger U = 1$ we see that

$$A^\dagger A = V^\dagger S^2 V \tag{2.90}$$

By construction, the matrix $A^\dagger A$ is Hermitian and positive semi-definite. Hence, we see that V forms its eigenvectors and σ_α^2 its eigenvalues. Similarly, using the fact that $VV^\dagger = 1$ we get

$$AA^\dagger = US^2U^\dagger \tag{2.91}$$

Thus, σ_α^2 are also the eigenvalues of $AA^\dagger$. It is interesting to note that when A is rectangular, $A^\dagger A$ and $AA^\dagger$ will have different dimensions. The point is that the largest of the two will have the same eigenvalues as the smaller one, plus a bunch of zero eigenvalues. The only type of matrix for which the singular values are identically equal to the eigenvalues are positive semi-definite matrices, like density matrices ρ.

One of the most important applications of the SVD is in making **low rank approximations** of matrices. To do that, suppose A is $N \times N$. Then it will have N^2 entries which, if N is large, will mean a bunch of entries. But now let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors of size N and consider the outer product $\boldsymbol{u}\boldsymbol{v}^\dagger$, which is also an $N \times N$ matrix with entries $(\boldsymbol{u}\boldsymbol{v}^\dagger)_{ij} = u_i v_j^*$. We then see that even though this is $N \times N$, the entries of this matrix are not independent, but are completely specified by the $2N$ numbers u_i and v_i. A matrix of this form is called a rank-1 matrix (just like the rank-1 projectors we studied before).

Going back now to Eq. (2.89), let $\boldsymbol{u}_\alpha$ denote a column vector with entries $U_{i\alpha}$ and, similarly, let $\boldsymbol{v}_\alpha$ denote a column vector with entries $V_{j\alpha}$. Then it is easy to verify that the matrix A in Eq. (2.89) can be written as

$$A = \sum_{\alpha=1}^{r} \sigma_\alpha \boldsymbol{u}_\alpha \boldsymbol{v}_\alpha^\dagger. \tag{2.92}$$

We have therefore decomposed the matrix A into a sum of rank-1 matrices, weighted by the singular values σ_α. Since the singular values are always non-negative and appear in decreasing order, we can now think about retaining only the largest singular values. That is, instead of summing over the full Schmidt rank r, we sum only up to a smaller number of singular values $r' < r$ to get an approximate representation of A:

$$A' = \sum_{\alpha=1}^{r'} \sigma_\alpha \boldsymbol{u}_\alpha \boldsymbol{v}_\alpha^\dagger.$$

This is called a rank-r' approximation for the matrix A. If we consider just the largest singular value (a rank-1 approximation) then we replaced N^2 elements by $2N$, which can be an enormous improvement if N is large. It turns out that this approximation is controllable in the sense that the matrix A' is the *best rank-r' approximation* of A given the Frobenius norm, defined as $||A|| = \sum_{ij} |A_{ij}|^2$. That is, A' is the rank-r' matrix which minimizes $||A - A'||$.

Schmidt decomposition

I have introduced above the SVD as a general matrix decomposition, which is useful to know since it appears often in many fields of research. Now I want to apply the SVD to extract properties of quantum states. Consider again a bipartite system described by the pure state

$$|\psi\rangle = \sum_{a,b} \psi_{ab} |a, b\rangle. \tag{2.93}$$

With a moment of though we see that ψ_{ab} can also be interpreted as a matrix of coefficients. In fact, this matrix will in general be rectangular when the dimensions d_A and d_B are different. We may then apply the SVD to the matrix with entries ψ_{ab}. Using Eq. (2.89) we see that this decomposition will have the form

$$\psi_{ab} = \sum_{\alpha} \sigma_\alpha U_{a\alpha} V_{b\alpha}^*. \tag{2.94}$$

The matrix ψ_{ab} is special in that the state $|\psi\rangle$ must be normalized. This means that $\sum_{ab} |\psi_{ab}|^2 = 1$ which in turn implies that

$$\sum_{\alpha=1}^{r} \sigma_\alpha^2 = 1. \tag{2.95}$$

In general the singular values are simply non-negative. But for states ψ_{ab} they are also normalized in this way.

Inserting Eq. (2.94) back into (2.93) now gives

$$|\psi\rangle = \sum_{a,b,\alpha} \sigma_\alpha U_{a\alpha} V^*_{b\alpha} |a,b\rangle = \sum_\alpha \sigma_\alpha \bigg[\sum_a U_{a\alpha}|a\rangle\bigg] \otimes \bigg[\sum_b V^*_{b\alpha}|b\rangle\bigg]. \tag{2.96}$$

We now define new sets of states for systems A and B, as

$$|\alpha_A\rangle = \sum_a U_{a\alpha}|a\rangle, \tag{2.97}$$

$$|\alpha_B\rangle = \sum_b V^*_{b\alpha}|b\rangle. \tag{2.98}$$

Note how these states are labeled by the same index α, even though they may be completely different (recall that we can even have $d_A \neq d_B$). Notwithstanding, we notice that these states are orthonormal because of the properties of the SVD matrices U and V.

Thus, we can now write our entangled state $|\psi\rangle$ as

$$\boxed{|\psi\rangle = \sum_\alpha \sigma_\alpha |\alpha_A\rangle \otimes |\alpha_B\rangle.} \tag{2.99}$$

This is way better than (2.93) because now we only have a single sum. It is a bit like we diagonalized something (but what we did was find the singular values of ψ_{ab}). Note also that this is exactly the type of state that we used in Eq. (2.59) when we first introduced the connection between mixed states and entanglement. The step in going from a general entangled state (2.93) to a state of the form (2.99) is called the **Schmidt decomposition** of the state. The square of the singular values, $\lambda_\alpha := \sigma_\alpha^2$, are also called Schmidt coefficients. As we will see, all the information about entanglement is contained in these guys.

We have seen that a general state such as (2.93) will be a product state when $\psi_{ab} = f_a g_b$ is a product of coefficients. But that can in practice be a hard thing to check. If we look at the Schmidt form (2.99), however, it is now trivial to know when the state will be a product or not: it will only be a product if $\sigma_1 = 1$ and all other $\sigma_\alpha = 0$. That is, they will be in a product state when the Schmidt rank is $r = 1$. We can even go further and use the singular values/Schmidt coefficients to quantify the the *degree* of entanglement. To do that, we compute the reduced density matrices of A and B, starting from the state (2.99). Since the states $|\alpha_A\rangle$ and $|\alpha_B\rangle$ are orthonormal, it is straightforward to find that

$$\rho_A = \sum_\alpha \sigma_\alpha^2 |\alpha_A\rangle\langle\alpha_A|, \tag{2.100}$$

$$\rho_B = \sum_\alpha \sigma_\alpha^2 |\alpha_B\rangle\langle\alpha_B|. \tag{2.101}$$

Once we have these reduced density matrices, we can now compute their purity:

$$\boxed{\operatorname{tr}(\rho_A^2) = \operatorname{tr}(\rho_B^2) = \sum_\alpha \sigma_\alpha^4 = \sum_\alpha \lambda_\alpha^2.} \tag{2.102}$$

Quite remarkably, we see that the purity of A and B are equal (which is true even if one has $d_A = 2$ and the other has $d_B = 1000$). Thus, we conclude that *the purity of the reduced states can be directly used as a quantifier of entanglement.* The more entangled are two systems, the more mixed are their reduced density matrices.

To summarize, I want to emphasize that *all* entanglement properties can be obtained from the singular values of ψ_{ab}. If one such singular value is $\sigma_1 = 1$ then the others must be zero so the two parties are in a product state. Otherwise, their degree of entanglement is quantified by the sum in Eq. (2.102). In particular, we now finally have the tools to define what a *maximally entangled state* is: a maximally entangled state is a state in which all singular values are equal. Due to the normalization (2.95), this then implies

$$\sigma_\alpha = \frac{1}{\sqrt{r}}. \tag{2.103}$$

As an example, consider a state of the form

$$|\psi\rangle = \cos(\theta/2)|0,1\rangle + \sin(\theta/2)|1,0\rangle. \tag{2.104}$$

We already know that if $\theta = 0, \pi$ the state will be a product and if $\theta = \pi/2$ the state will be a Bell state (2.50). In this case the matrix ψ_{ab} has coefficients $\psi_{01} = \cos(\theta/2)$ and $\psi_{10} = \sin(\theta/2)$. The singular values can be found either by asking semiconductors to do it for you or by computing the eigenvalues of $\psi^\dagger\psi$. In either case, they are

$$\sigma_1 = \cos(\theta/2), \qquad \sigma_2 = \sin(\theta/2). \tag{2.105}$$

Thus we see that when $\theta = 0, \pi$ we have one of the singular values equal to 1, which is the case of a product state. Conversely, we see that the singular values will be all equal when $\theta = \pi/2$. Thus, the Bell state (2.50) is the maximally entangled state.

So far we have only considered the entanglement between bipartite systems which are in *pure* states. A natural question therefore is how to quantify the degree of entanglement between parties that are in mixed states. That is, when not only are ρ_A and ρ_B mixed, but when ρ_{AB} itself is already mixed. This question is actually much harder and is still an open topic of research. The reason is that it is not easy to distinguish between *quantum correlations* and *classical correlations*. To see what I mean, have a look back at the state (2.81). This is a classical probability distribution. However, the sub-systems A and B are not statistically independent because $p_{i,j}$ cannot be factored as a product of two probability distributions. This is therefore an instance of classical correlations. For more general states, it is not easy to separate it from true quantum features. In fact, in this case there is even more than one type of quantum correlation (for instance, a famous one is the so called *quantum discord*). We will get back to this topic later in the course.

Multipartite entanglement

The quantification of entanglement in a multipartite system A, B, C, ... is a difficult task and still an open research problem. One thing that can be done, though, is to look at the entanglement of all bipartitions. To see how that works, consider again a general N-partite state

$$|\psi\rangle = \sum_{s_1,\ldots,s_N} \psi_{s_1\ldots s_N} |s_1,\ldots,s_N\rangle. \tag{2.106}$$

The key now is to try to map this into the problem we just discussed, which can be done using the idea of *collective indices*. For instance, suppose we want to make a bipartition such as $1,\ldots,k$ and $k+1,\ldots,N$. Then we define two collective indeces

$$a = \{s_1,\ldots,s_k\}, \qquad b = \{s_{k+1},\ldots,s_N\} \tag{2.107}$$

so that the state (2.106) is now mapped back into state (2.93). We can then use the usual Schmidt procedure we just described.

This idea of collective indices is really important and really abstract at first. The point to remember is that this is only a *relabelling* of stuff. For instance, suppose we have two qubits with states $s_i = \{0,1\}$. Then we can define a collective index by means of a correspondence table. For instance we can say $(0,0)$ is $a = 1$,$(0,1)$ is $a = 2$ and so on. We usually write this symbolically as follows:

$$\psi_{s_1\ldots s_N} = \psi_{(s_1\ldots s_k),(s_{k+1}\ldots s_N)}. \tag{2.108}$$

This means we have grouped the big tensor into two blocks and now it behaves as matrix with only two collective indices. This type of operation is really annoying to do by hand, but computationally it is not hard since it is simply a matter of relabelling.

State purification

We finish this section with the concept of purifying a state. Consider a physical system A described by a general mixed state ρ_A with diagonal form

$$\rho = \sum_a p_a |a\rangle\langle a|$$

By purification we mean writing this mixed state as a pure state in a larger Hilbert space. Of course, this can obviously always be done because we can always imagine that A is mixed because it was entangled with some other system B. All we need is to make that formal. One thing we note from the start is that this operation is certainly not unique since the system B can have any size. Thus, there is an infinite number of pure states which purify ρ_A. The simplest approach is then to consider B to be a copy of A. We then define the pure state

$$|\psi\rangle = \sum_a \sqrt{p_a}|a\rangle \otimes |a\rangle \tag{2.109}$$

Tracing over B we get

$$\mathrm{tr}_R\,|\psi\rangle\langle\psi| = \rho \tag{2.110}$$

Thus, $|\psi\rangle$ is a purified version of ρ, which lives in a doubled Hilbert space. Notice how the probabilities p_a appear naturally here as the Schmidt coefficients.

2.9 Entropy and mutual information

The concept of entropy plays a central role in classical and quantum information theory. In its simplest interpretation, entropy is a measure of the disorder (or mixedness) of a density matrix, a bit like the purity $\mathrm{tr}(\rho^2)$. But with entropy this disorder acquires a more informational sense. We will therefore start to associate entropy with questions like "how much information is stored in my system". Also like the purity, entropy can be used to quantify the degree of correlation between systems. And that makes sense because correlation is a measure of information: when two systems are correlated we can ask questions such as "how much information about A is stored in B". Unlike the purity, however, entropy will also serve to quantify correlations of mixed states, which is done using the concept of **mutual information**. We will also introduce another concept called **relative entropy** which plays the role of a "distance" between two density matrices. It turns out that the relative entropy is not only useful in itself, but it is also useful as a tool to prove certain mathematical identities.

In thermodynamics we like to associate entropy with a unique physical quantity. In quantum information theory that is not exactly the case. There is one entropy, called the **von Neumann entropy**, which does have a prominent role. However, there are also other entropy measures which are also of relevance. An important family of such functions are the so-called **Rényi entropies**, which contain the von Neumann entropy as a particular case. We will also discuss them a bit.

The von Neumann entropy

Given a density matrix ρ, the von Neumann entropy is defined as

$$\boxed{S(\rho) = -\,\mathrm{tr}(\rho \ln \rho) = -\sum_k p_k \ln p_k.} \tag{2.111}$$