

Lecture 4: Quantifying entanglement.

October 2nd 2026

In the previous lecture, we have defined entanglement as a black-and-white notion separating product (or factorized) states on the one hand, and entangled states on the other hand.

MCQ: which is the most entangled state among the two states below?

However, even if state

$$|\Psi_1\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

and state

$$|\Psi_2\rangle = \sqrt{0.1}|00\rangle + \sqrt{0.9}|11\rangle$$

are both entangled states, it is obvious that the second state is less entangled than the first one. In this lecture, we are thus going to introduce metrics to *quantify* entanglement. For this, we will rely heavily on a graphical representation of states and operations on states called *tensor networks*.

1 A graphical representation of quantum states: tensors

The goal of the graphical representation we are about to introduce is (among others) to make entanglement (or lack thereof) a visual notion. For this, we are going to represent vector, matrices and higher-dimensional linear-algebra objects as diagrams.

1.1 Penrose diagrams for tensors

1.1.1 Blob notation

We are going to represent multidimensional arrays (which we will call “tensors”) as blobs with a number of “legs” equal to the number of dimensions (also called “rank”¹) of the array. For instance, vectors are 1-dimensional arrays v_i , matrices 2-dimensional arrays A_{ij} . 0-dimensional arrays are scalars. The corresponding “Penrose diagrams” are shown in Fig. 1.

An interesting special case of matrix is the identity matrix, which also corresponds to Kronecker delta:

$$I_{ij} = \delta_{ij}.$$

Since by definition, such an operator imposes the output index to be equal to the input index, the graphical representation for it is a line, see Fig. 1 (a).

¹not to be confused with the rank of a matrix, defined as the number of independent columns or rows of a matrix

1.1.2 Index grouping

The legs (indices) of a tensor can be grouped. For instance, a matrix A_{ij} can be reinterpreted as a vector $A_{\alpha=(ij)}$. This is illustrated in Fig. 1 (b). Take the example of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

It can equivalently be represented by $A = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$. The new index α is given as a function

of the old ones (i, j) as $\alpha = Mi + j$ if A is a $N \times M$ matrix. The reverse operation is of course possible, by recovering i and j as the quotient and remainder of the Euclidean division of α by M . This particular way of going from a matrix to a vector is called (row-major) vectorization. We could also have stacked the columns instead of the row (column-major vectorization). Note that in our Penrose convention, there is no telling one vectorization from the other.

We can now turn to a more quantum mechanical example. The wave function of a system of n qubits reads, in the “computational basis”:

$$|\Psi\rangle = \sum_{b_1, b_2 \dots b_n, b_i \in \{0,1\}} \psi_{b_1, b_2 \dots b_n} |b_1, b_2 \dots b_n\rangle. \quad (1)$$

The corresponding Penrose diagram is shown in Fig 1 (b). It can alternatively but equivalently be written as

$$|\Psi\rangle = \sum_{i=0}^{2^n-1} \psi_i |i\rangle.$$

This corresponds to grouping the n legs of dimension 2 to 1 leg of dimension 2^n . See Fig. 1 (b). We see that the legs of a tensor represent graphically the tensor product structure of the Hilbert space. Grouping them corresponds to changing the tensor product structure.

1.1.3 Trace

The trace of a matrix corresponds to the operation

$$\text{Tr}(A) = \sum_i A_{ii}.$$

Graphically, this is represented as in Fig. 1 (c): the indices that are summed over are linked with a line. This way, the diagram has no open (also called “external”) lines anymore: it represents a 0-dimensional array, namely a scalar, as expected.

A generalization of the trace, called partial trace, can be defined graphically by summing over some indices, but not all. For instance, for an operator A acting on a two-qubit Hilbert space, the matrix elements look like $A_{b_1 b_2, b'_1 b'_2}$. Partial tracing over the second Hilbert space corresponds to the diagram shown in Fig. 1 (c) and the formula

$$[\text{Tr}_{\mathcal{H}_2} A]_{b_1, b'_1} = \sum_{b_2} A_{b_1 b_2, b'_1 b_2}. \quad (2)$$

Q: Compute the partial trace corresponding to state $|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$, namely the partial trace (on e.g the second qubit) of matrix

$$|\psi\rangle\langle\psi| = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

What about $|\chi\rangle = \frac{|00\rangle + |01\rangle + |10\rangle + |11\rangle}{2}$?

1.1.4 Computing storage and runtime complexity

A powerful feature of this graphical notation is that it makes it easy to compute the cost of storing a given object in the memory of a classical computer. This cost is the product of the dimensions of its legs. For instance, if the matrix A shown in 1 (a) has indices i and j of respective dimensions M and N , then storing it costs $M \times N$ complex numbers. Of course, this cost should not depend on the grouping of the indices: one can check that grouping two indices of dimensions M and N yields one index of dimension $M \times N$.

A telling example is the wavefunction $|\Psi\rangle$ we introduced above: it costs $2 \times 2 \times \dots \times 2 = 2^n$ complex numbers to store because it is represented by n legs $b_1, \dots, b_n$ of dimension 2 (or one leg i of dimension 2^n). Of course, grouping does not change storage cost.

In addition to storage, the cost of performing operations (runtime complexity) on the tensors can easily be computed. Let us take the partial trace as an example: the cost of performing this operation is given by $2 \times 2 \times 2$: for all (b_1, b'_1) pairs, we sum over b_2 (see Eq. (2)). Graphically (Fig. 1 (c)), we see that the Penrose diagram describing the operation has three legs of dimension 2 (two external legs corresponding to b_1, b'_1 and one internal one, b_2 , corresponding to the partial tracing), hence a cost 2^3 .

1.2 Entanglement (or absence thereof) in tensor notation

Let us assume our wavefunction factorizes between two portions of our total Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$:

$$|\Psi\rangle = |\varphi_A\rangle \otimes |\chi_B\rangle$$

with $|\varphi_A\rangle \in \mathcal{H}_A$ and $|\chi_B\rangle \in \mathcal{H}_B$. Then, decomposing both factors on the respective bases $\{|b_1, \dots, b_m\rangle\}$ and $\{|b_{m+1}, \dots, b_n\rangle\}$ of $\mathcal{H}_A$ and $\mathcal{H}_B$:

$$|\varphi_A\rangle = \sum_{b_1, \dots, b_m} \varphi_{b_1, \dots, b_m} |b_1, \dots, b_m\rangle, \quad |\chi_B\rangle = \sum_{b_{m+1}, \dots, b_n} \chi_{b_{m+1}, \dots, b_n} |b_{m+1}, \dots, b_n\rangle,$$

we obtain:

$$|\Psi\rangle = \sum_{b_1 \dots b_n} \psi_{b_1 \dots b_n} |b_1 \dots b_n\rangle$$

with

$$\psi_{b_1 \dots b_n} = \varphi_{b_1, \dots, b_m} \chi_{b_{m+1}, \dots, b_n}. \quad (3)$$

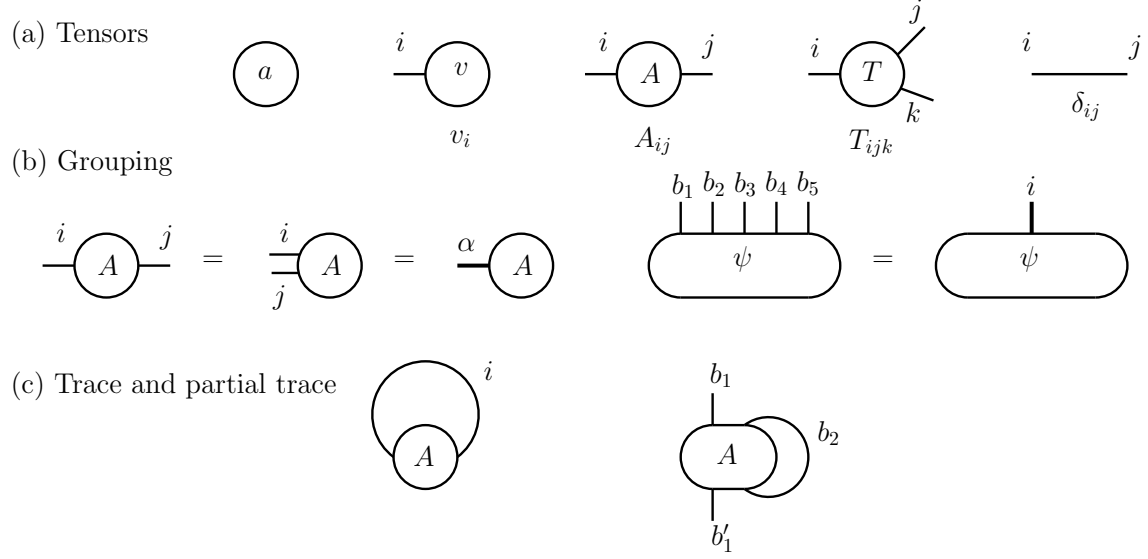


Figure 1: Tensors: illustrations.

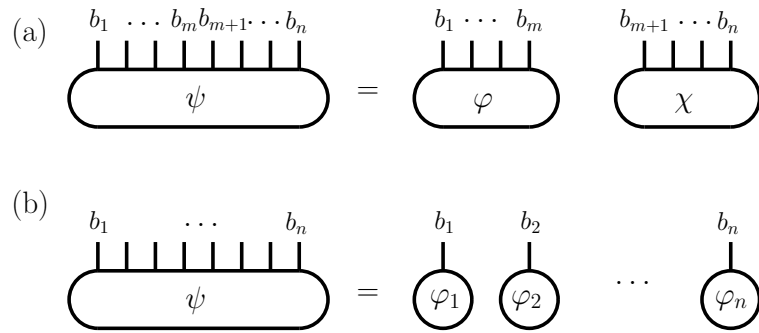


Figure 2: Graphical representation of a factorized state in two (panel (a)) or more states (panel (b)).

To represent the right-hand side of this equality graphically, we just use two disconnected blobs, as shown in 2 (a). With this representation, a wavefunction that factorizes between two Hilbert spaces is represented by a diagram that splits into two disconnected diagrams, one for each Hilbert space.

MCQ: What is the cost of storing wavefunction that is factorized between two equal subsystems?

We can use this representation to compute the storage cost of such a wavefunction. For simplicity, let us assume we are looking at an equal partition, $m = n/2$. Then, to store Ψ , we only need to store the coefficients of $|\psi_A\rangle$ and $|\chi_B\rangle$ individually: this totals to $2 \times 2^{n/2}$ complex numbers. Graphically, we just compute the storage of each of the two blobs ($2^{n/2}$ because both have $n/2$ legs) and sum them. We thus see that storing a product state ($2^{n/2+1}$ complex numbers) is much less expensive than storing a fully entangled state (2^n complex numbers).

MCQ: What is the cost of storing a wavefunction that is factorized between all qubits?

We can extend this to more than two Hilbert spaces, namely to the notion of multipartite (as opposed to bipartite) entanglement: what if the qubits are not entangled at all? Then, the wavefunction factorizes as

$$|\Psi\rangle = |\varphi_1\rangle \otimes |\varphi_2\rangle \otimes \cdots \otimes |\varphi_n\rangle,$$

and thus

$$\psi_{b_1 \dots b_n} = (\varphi_1)_{b_1} (\varphi_2)_{b_2} \cdots (\varphi_n)_{b_n}, \quad (4)$$

which is represented in Fig. 2 (b). The storage cost of such a fully factorized wavefunction is $2 \times n$ (since each tensor φ_i has one leg of size 2). This drastic reduction of storage complexity (from 2^n to $2n$!) makes it obvious why entanglement is a necessary ingredient to make quantum computing powerful: in the absence of it, a classical computer could easily manipulate the states since it would need to store only $2n$ complex numbers (not even counting normalization).

Yet, while this representation makes it easy to distinguish between product states and entangled states, we are still missing a way to quantitatively distinguish two entangled states like the $|\Psi_1\rangle$ and $|\Psi_2\rangle$ states above. Intuitively, we can anticipate that it has to do with the ability to “split” diagrams in two: while a wavefunction diagram that can be completely split in two disconnected pieces has no entanglement, there is a way to quantify the amount of connection between two parts of a diagram. This is the topic of the next section.

2 Quantifying entanglement

2.1 Tensor networks: combining tensors

MCQ: what does diagram (a) mean?

We have so far learned how to represent simple algebraic objects like vectors, matrices, etc. *Tensor networks* is a graphical representation of operations between such objects. For

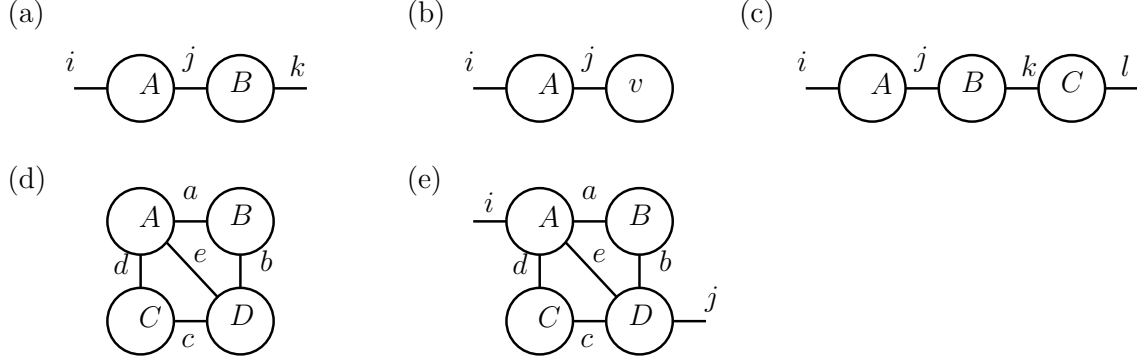


Figure 3: Examples of tensor networks.

instance, a matrix-matrix multiplication,

$$C_{ik} = \sum_j A_{ij} B_{jk} \quad (5)$$

is represented by linking the two blobs representing A and B by an edge representing the internal summation j , as shown in Fig. 3 (a). Like the partial trace example, the cost of the operation is readily computed by looking at the tensor network: one just multiplies the dimensions of the legs: here, assuming legs of dimension N , since we have three legs (two external, i and k , and one internal, j), the cost is N^3 . One can check that it is also the cost one would have found by inspecting Eq. (5).

MCQ: what is the runtime complexity of diagram (c)?

In this Figure 3, we also see other common operations:

1. Matrix vector multiplication. (Fig. 3 (b)). *Cost?*
2. Multiplying three matrices ABC . (Fig. 3 (c)). *Cost? As it turns out, there are several ways to perform this computation. One can just directly perform a summation over the internal indices:*

$$D_{il} = \sum_{jk} A_{ij} B_{jk} C_{kl}$$

which gives a complexity N^4 (two internal + two external legs). One can as well first compute the product $B' = BC$ (N^3) and then AB' (N^3), resulting in a N^3 complexity.

More formally, a tensor network is graph $G = (V, E)$ (with V and E the vertex and edge set, respectively) where each vertex represents a tensor, and edges represent internal summations over shared indices of two tensors. This may lead to more complicated examples than the ones we are used to. For instance, in Figure 3 (d)-(e), we display two examples of more sophisticated tensor networks:

1. Four tensors in a square, Fig. 3 (d): means $\sum_{abcde} \dots$
2. Same example with “external legs”: (Fig. 3 (e))

In fact, applying a gate to a quantum state is also a tensor network. For instance, applying a Hadamard gate to the first qubit of a n -qubit wavefunction is represented by the tensor network shown in Fig. 4. As for the other examples, we can directly assess the complexity

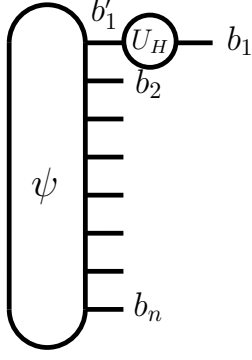


Figure 4: Gate application as a tensor network.

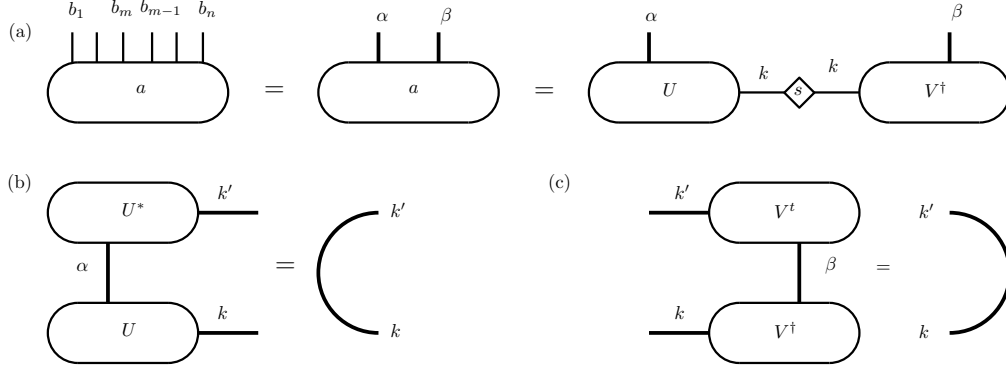


Figure 5: Singular value decomposition and MPS. (a) Singular value decomposition and (b-c) orthogonality properties of the corresponding matrices.

of such a computation: 2^{n+1} (n external legs and 1 internal leg, all of dimension 2). This is cheaper than a naive application of the gate as a matrix-vector $U|\psi\rangle$ multiplication, which would cost $2^n \times 2^n = 4^n$.

2.2 Splitting tensors: singular value decomposition

Now that we know how to combine tensors, let us learn how to split them. We already came across an extreme case of splitting: when a wavefunction is factorizable, it can be completely disconnected in two (or more parts), as illustrated in Fig. 2.

The central mathematical tool to find factorized forms of tensors in the singular value decomposition (SVD). The SVD of a $M \times N$ matrix a reads:

$$a_{\alpha,\beta} = \sum_{k=0}^{r-1} U_{\alpha,k} s_k V_{k,\beta}^\dagger, \quad (6)$$

with U a $M \times r$ matrix, $V^\dagger$ a $r \times N$ matrix and $U^\dagger U = I$, $V^\dagger V = I$ (U and V are sometimes called semi-unitary matrices). This decomposition is illustrated in a graphical manner in Fig. 5 (a), and the properties of U and V are graphically depicted in Fig. 5 (b).

The s_k 's are real numbers called the singular values of a . There are r nonzero singular values, where $r \leq \min(M, N)$. In fact, r can be shown to be the rank of matrix a .

2.3 A special case: the Schmidt decomposition, and von Neumann entanglement entropy

The SVD can be applied to a wavefunction. For this, given a bipartition of the system into a subspace A (say the first m qubits) and a subspace B (say the other $n - m$ qubits), we group the indices of the wavefunction $\psi_{b_1, \dots, b_n}$ into two compound indices $\alpha = (b_1, \dots, b_m)$ and $\beta = (b_{m+1}, \dots, b_n)$. The wavefunction (1) thus becomes

$$|\Psi\rangle = \sum_{\alpha=0}^{2^m-1} \sum_{\beta=0}^{2^{n-m}-1} \psi_{\alpha,\beta} |\alpha\rangle \otimes |\beta\rangle. \quad (7)$$

The coefficients $\psi_{\alpha,\beta}$ can now be interpreted as the coefficients of a $2^m \times 2^{n-m}$ matrix ψ , to which we now apply a SVD (Eq. (6)). Plugging this decomposition into (7), we get, after reshuffling the three summation symbols,

$$|\Psi\rangle = \sum_{k=0}^{r-1} s_k \left(\sum_{\alpha=0}^{2^m-1} U_{\alpha,k} |\alpha\rangle \right) \left(\sum_{\beta=0}^{2^{n-m}-1} V_{k,\beta}^\dagger |\beta\rangle \right).$$

Defining $|\varphi_k\rangle = \sum_{\alpha=0}^{2^m-1} U_{\alpha,k} |\alpha\rangle$ and $|\phi_k\rangle = \sum_{\beta=0}^{2^{n-m}-1} V_{k,\beta}^\dagger |\beta\rangle$, we obtain the so-called Schmidt decomposition of the state:

$$|\Psi\rangle = \sum_{k=0}^{r-1} s_k |\varphi_k\rangle |\phi_k\rangle, \quad (8)$$

i.e a decomposition of the state as a sum of tensor products of states living in A and B , respectively. Importantly, thanks to the semi-unitarity of U and V , the $\{|\varphi_k\rangle\}_k$ and $\{|\phi_k\rangle\}_k$ form two orthonormal bases. Using this property, we have:

$$\langle\Psi|\Psi\rangle = \sum_k s_k^2 = 1, \quad (9)$$

due to the normalization of $|\Psi\rangle$. This can also be checked graphically.

We note that if $r = 1$, then $s_0 = 1$ and thus

$$|\Psi\rangle = |\varphi_0\rangle |\phi_0\rangle$$

namely it is not entangled. If $r > 1$, by definition, $|\Psi\rangle$ is entangled.

MCQ: Which are valid Schmidt decompositions?

(a) $|\psi\rangle = \frac{|0,0\rangle + |1,1\rangle}{\sqrt{2}}$

(b) $|\psi\rangle = \frac{|+,0\rangle + |-,1\rangle}{\sqrt{2}}$

(c) $|\psi\rangle = \frac{1}{\sqrt{3}}|0,+\rangle + \sqrt{\frac{2}{3}}|-,1\rangle$

(d) $|\psi\rangle = \frac{|0,0\rangle + |+,1\rangle}{\sqrt{2}}$

Perform the SVD of state (d).

In fact, the singular values can be used to give a more quantitative appraisal of the level of entanglement of a state. It follows from (9) that $0 \leq s_k^2 \leq 1$ and thus the $\{s_k^2\}_k$ define a valid

(discrete) probability distribution. We can thus define the Shannon entropy associated with this distribution:

$$S_{AB} = - \sum_k s_k^2 \log_2 s_k^2. \quad (10)$$

We put a subscript AB to make the dependence on the initial bipartition explicit (but we will often drop the explicit mention of the bipartition). This quantity is called the *von Neumann entanglement entropy*.

Let us examine its properties:

- It is minimal when the distribution of the squared singular values is peaked ($s_0 = 1$, $s_{k>0} = 0$). In this case, namely when the state is not entangled, $S_{AB} = 0$.
- It is maximal when the information content of the distribution is the smallest, namely when the singular values are constant $s_k = \text{const}$. Since $r \leq \min(2^m, 2^{n-m})$, the maximum rank can be attained for an equal bipartition ($m = n - m = n/2$), in which case, using normalization, we can compute explicitly the value of the constant: $s_k = 1/\sqrt{2^{n/2}}$. Thus,

$$S_{\max} = -2^{n/2} 1/2^{n/2} (-n/2) = \frac{n}{2} \quad (11)$$

i.e the maximal entanglement entropy for the bipartition of a n -qubit Hilbert space is $n/2$. A state attaining this maximum is said to be maximally entangled, and is said have a volume-law entanglement (where the volume here denotes the number n of qubits).

2.4 Entanglement entropy of interesting states

We are now in a position to quantify the entanglement of some states of interest. Let us start with our preliminary examples, $|\Psi_1\rangle$ and $|\Psi_2\rangle$ above. Since $\{|0\rangle, |1\rangle\}$ forms an orthonormal basis, both states are already in a Schmidt form. From their singular values $\{1/\sqrt{2}, 1/\sqrt{2}\}$ and $\{\sqrt{0.1}, \sqrt{0.9}\}$, respectively, we can compute the entanglement entropy for the bipartition between qubit 1 and qubit 2:

$$S_1 = n/2 = 2/2 = 1$$

$$S_2 = -0.1 \log_2 0.1 - 0.9 \log_2 0.9 \approx 0.47.$$

For the first computation, we directly used the fact that constant singular values will maximize S . We thus see that, as expected, the second state is less entangled than the first one.

Q: Compute the von Neumann entanglement entropy of

$$|\Psi_2\rangle = \sqrt{\epsilon}|00\rangle + \sqrt{1-\epsilon}|11\rangle$$

MCQ: Is state

$$\frac{|00 \dots 0\rangle + |11 \dots 1\rangle}{\sqrt{2}}$$

maximally entangled?

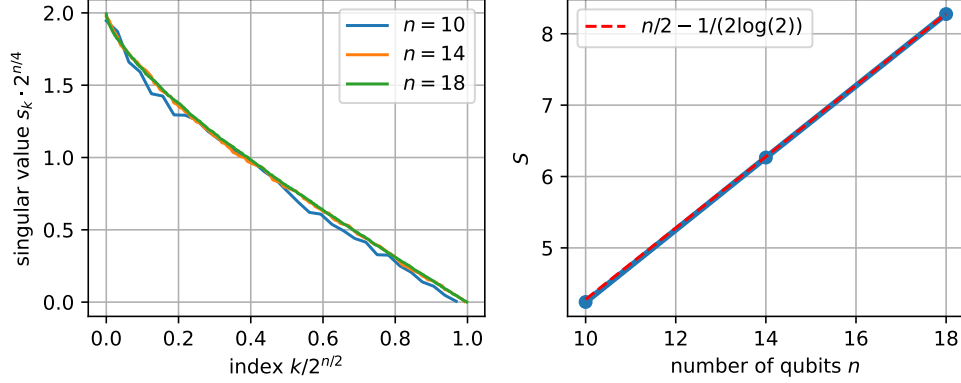


Figure 6: Schmidt values s_k as a function of k for different numbers of qubits n (left) and entanglement entropy S as a function of n (right) for random quantum states.

What about a generalized Bell state (called GHZ state for Greenberger, Horne and Zeilinger ([Greenberger *et al.* \(1989\)](#))))

$$|\Psi_3\rangle = \frac{|00 \dots 0\rangle + |11 \dots 1\rangle}{\sqrt{2}}?$$

Picking an equal bipartition (assuming an even number of qubits), we see that this state is already in Schmidt form with orthonormal bases $\{|b_1 \dots b_{n/2}\rangle, b_i \in \{0, 1\}\}$ and $\{|b_{n/2+1} \dots b_n\rangle, b_i \in \{0, 1\}\}$ for both subspaces and singular values $\underbrace{\{1/\sqrt{2}, 0, 0, \dots, 0, 1/\sqrt{2}\}}_{2^{n/2} \text{ values}}$. The corresponding

entanglement entropy is

$$S_3 = 1.$$

One sometimes says that a GHZ state has 1 ebit (for entanglement bit) of entanglement. It is not maximally entangled, since $S_3 = 1 < n/2$. It is thus natural to ask how to create maximally entangled states.

As it turns out, random quantum states are maximally entangled (asymptotically, namely in the large- n limit). One can be checked numerically (see problem set below) that these states have volume-law entanglement. This fact has been used extensively in recent years to design experiments to outperform classical computers: in these experiments, quantum programs were designed to produce random quantum states, with the intuition that because these states are maximally entangled, they would be difficult to store on classical computers.

2.5 Beyond bipartite entanglement: matrix product states and the area law

Instead of a simple bipartition, one can choose to split the original rank- n tensor ψ (that appears in Eq. (7)) in n rank-2 tensors by successive SVD decompositions. This is illustrated with a graphical language in Fig. 5 (a), where a first bipartition between the first qubit and the rest of the qubits is done, the singular values s_1 and the $V_1^\dagger$ tensor are grouped into a

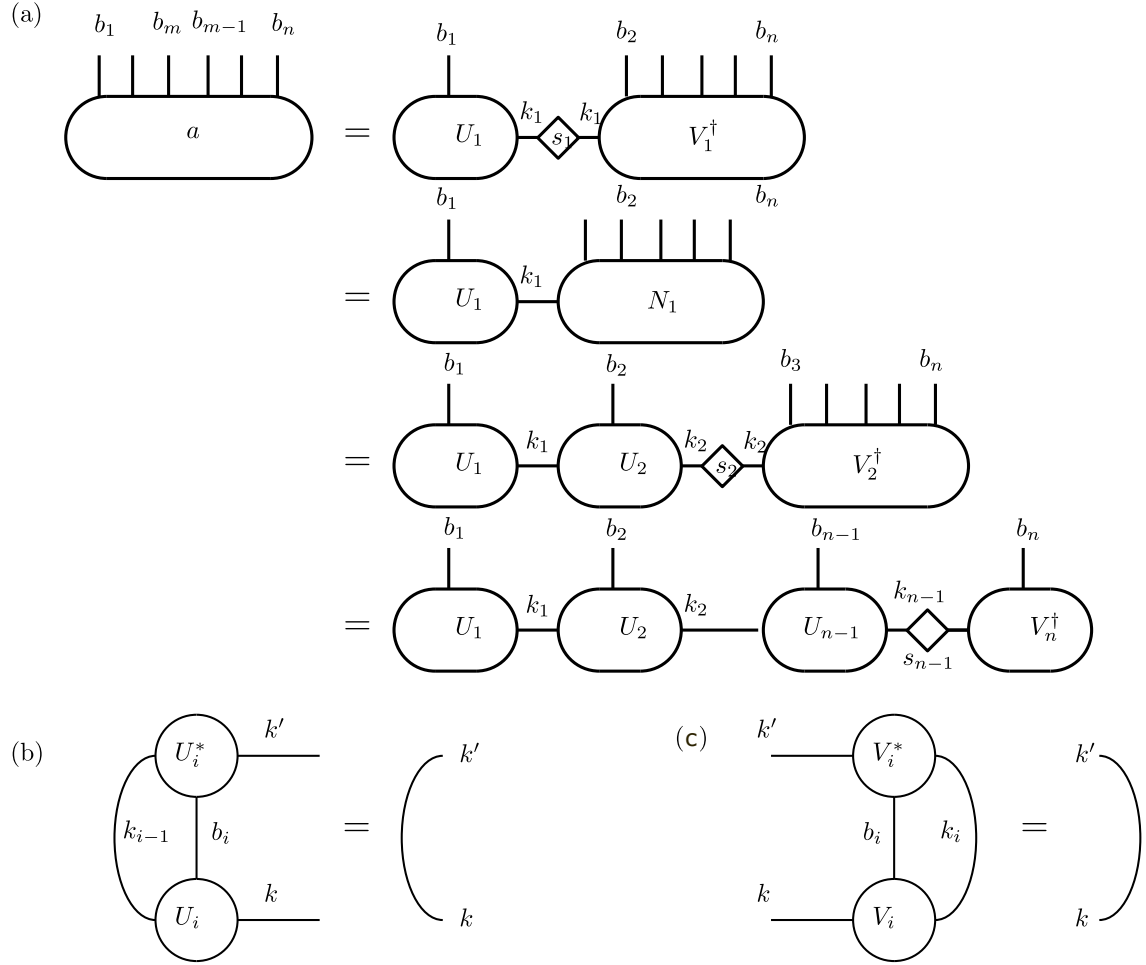


Figure 7: Matrix product states. (a) Construction of a matrix product state by successive SVDs. (b) Left-orthogonality conditions. (c) Right-orthogonality condition.

tensor N_1 which is then split between the second qubit and the rest by SVD, and so on and so forth until one obtains the following decomposition

$$\psi_{b_1 \dots b_n} = \sum_{k_1 \dots k_{n-1}} [U_1]_{k_1}^{b_1} [U_2]_{k_1, k_2}^{b_2} \cdots [U_{n-1}]_{k_{n-2}, k_{n-1}}^{b_{n-1}} [V_n^\dagger]_{k_{n-1}}^{b_n}, \quad (12)$$

namely the original tensor has been decomposed a product of matrices $U_1, U_2, \dots$ (we call them matrices because we consider the “external” indices b_i to be fixed). It is thus called a matrix product state (MPS), in analogy with a product state (see Eq. 4).

The dimension of the internal indices k_i is determined by the rank of the matrix that is split by SVD. Keeping in mind that the rank is the minimum of the two dimensions of the matrices, one can see that $k_1 \leq 2$, $k_2 \leq 4$ until $k_{n/2} \leq 2^{n/2}$ and then the maximal rank decreases to $k_{n-1} \leq 2$. Thus, $k_i \leq 2^{n/2}$ (for all i) if one follows the above exact decomposition.

In practice one seldom constructs an MPS from a full wavefunction since this would require storing this wavefunction in the first stages of the algorithm, which becomes impossible even for moderate n ’s. Rather, one often uses MPS in the context of the variational method, namely given a Hamiltonian $\hat{H}$ whose ground state we want to find, we minimize

$$E(\{U_1, U_2, \dots, V_n\}) = \langle \psi(\{U_1, U_2, \dots, V_n\}) | \hat{H} | \psi(\{U_1, U_2, \dots, V_n\}) \rangle$$

with $|\psi(\{U_1, U_2, \dots, V_n\})\rangle$ a MPS state whose tensors $\{U_1, U_2, \dots, V_n\}$ we wish to find by minimization (the variational method with MPS as a variational ansatz is called the “density-matrix renormalization group” (DMRG) method (White (1992))). For such a minimization, we cap the maximum “bond dimensions” of these tensors to a number χ , namely we impose the tensors to be at most of dimension $\chi \times 2 \times \chi$ (where 2 is the dimension of the b_i leg). Then, the singular values in the Schmidt decomposition of the resulting state will be of the form $s_0, s_1, \dots, s_{\chi-1}, 0, 0 \dots 0$ (since capping the bond dimension amounts to capping the maximal rank). This in turns implies an upper bound to the entanglement entropy of the corresponding MPS: indeed, for a fixed number χ of nonzero singular values, the maximum entanglement entropy is reached when all the nonzero singular values are equal. Because the state is normalized, this equal value is $1/\sqrt{\chi}$ and thus

$$S_{\max}(\chi) = -\chi \frac{1}{\chi} \log_2(1/\chi) = \log_2(\chi).$$

In other words

$$S_{\text{MPS}} \leq \log_2(\chi). \quad (13)$$

The entanglement entropy of an MPS of maximal bond dimension χ is at most $\log_2(\chi)$. Conversely, if one wants to store a state with entanglement entropy S as an MPS, one needs its bond dimension to be larger or equal to 2^S . The total storage cost is the sum of the individual storage costs of each tensor, namely $2\chi^2$, which gives a total of $2n\chi^2$. If $\chi \sim 2^S$, we have a storage in $O(n4^S)$... instead of the naive 2^n storage cost that we started with before knowing about the notion of entanglement entropy!

This representation is thus very beneficial when S is small! Surprisingly, this happens more often than not in the physical world. For instance, ground states of local Hamiltonians

typically have a reasonably small entanglement entropy. More precisely, some of them obey a so-called “area law”, namely their entanglement entropy is proportional to the area (rather than the volume) of the system:

- For 2D systems, obeying an area law means that the entropy on a $n = L \times L$ grid will scale as $S \propto L = \sqrt{n}$, and hence a scaling of the MPS bond dimension of $\chi \leq 2^{\alpha\sqrt{n}}$, which is not polynomial but still subexponential.
- For 1D systems, obeying an area law leads to an even more drastic behavior of S : $S = O(1)$, namely it is independent on system size. In practice, it means that MPS can capture area-law states in 1D for huge systems.

From a formal point of view, ground states of 1D, gapped Hamiltonians have been proven to obey an area law ([Hastings \(2007\)](#)), while for 2D gapped systems it is only a conjecture.

3 The density matrix

The definition of the entanglement entropy we used above is based on the singular values obtained by performing a SVD of the amplitudes of the state of the system. In practice, we seldom have access to these amplitudes. Yet we may have access to the density matrix ρ_A (introduced in the preliminary lecture by David Clément) of a subsystem: this density matrix can be used to define the entropy and then compute it.

3.1 The density matrix as a partial trace over the environment

This section is a reminder of the notion of density matrix. We suppose our system to be made up of qubits. It is entangled with the outside world, which we call the environment. Later, we will apply this construction to the case we studied above, namely the “system” will be subpart A, while the environment will be subpart B of our full Hilbert space, and we will consider the density matrix ρ_A of subsystem A. But for now A is our qubit(s) and B our environment.

If we take into account the existence of the environment, we must describe the full state of the system by a wavefunction $|\Psi_{\text{tot}}\rangle$ that, generically, is not a product state $|\Psi_{\text{qubits}}\rangle \otimes |\Psi_{\text{environment}}\rangle$ (see Fig. 8 (a)). Therefore, the mere concept of a “wavefunction of the qubits”, $|\Psi_{\text{qubits}}\rangle$ —which we have been using extensively so far—does not make sense anymore. However, making measurements on the qubits only still makes sense (see Fig. 8 (b)):

$$\langle O \rangle = \langle \Psi_{\text{tot}} | \hat{O} \otimes I | \Psi_{\text{tot}} \rangle. \quad (14)$$

This formal expression is correct, but useless in practice because, almost by definition, the environment is huge and therefore $|\Psi_{\text{tot}}\rangle$ cannot be measured, let alone represented. One may however wonder about the existence of a quantity that completely characterizes the state of the qubits, i.e a quantity from which any observable on the system can be measured, yet without storing all the details of the environment.

It turns out that such a quantity can be defined in the following way:

$$\rho = \text{Tr}_{\text{env}} |\Psi_{\text{tot}}\rangle \langle \Psi_{\text{tot}}| \quad (15)$$

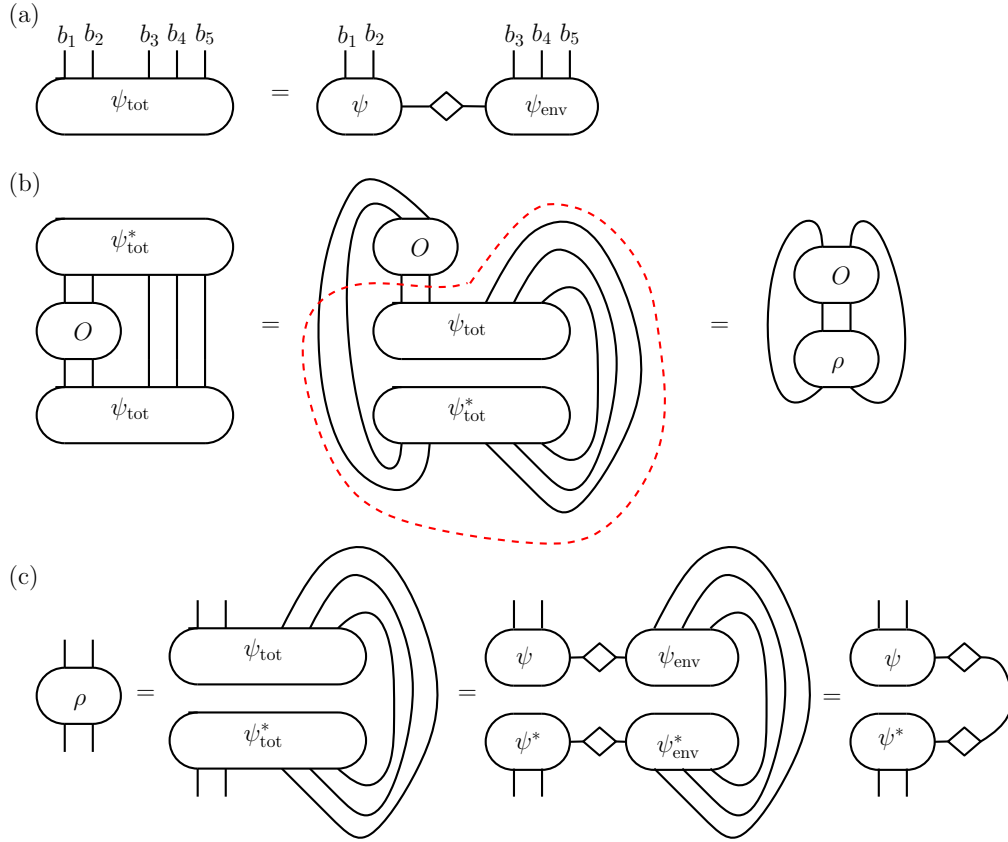


Figure 8: Graphical definition(s) of the density matrix. (a) System + environment wavefunction ψ_{tot} and its Schmidt decomposition (SVD). (b) Graphical representation of the average value $\langle O \rangle$ of an observable acting only on the system and its expression in terms of the density matrix. (c) Definition of the density matrix as a partial trace over the environment and, via the Schmidt decomposition of ψ_{tot} , as a statistical mixture of pure states.

where Tr_{env} denotes the “partial trace” over the environment. This definition is very natural, as shown graphically in Fig. 8 (b). Mathematically, if $|\varphi_k\rangle$ denotes a basis of the environment, then

$$\text{Tr}_{\text{env}} [\dots] = \sum_k \langle \varphi_k | \dots | \varphi_k \rangle.$$

ρ is called the density matrix of the qubits. We can compute its trace:

$$\begin{aligned} \text{Tr}(\rho) &= \text{Tr}_{\text{qubits}} (\text{Tr}_{\text{env}} |\Psi_{\text{tot}}\rangle \langle \Psi_{\text{tot}}|) \\ &= \text{Tr}_{\text{tot}} (|\Psi_{\text{tot}}\rangle \langle \Psi_{\text{tot}}|) \\ &= \langle \Psi_{\text{tot}} | \Psi_{\text{tot}} \rangle \\ &= 1. \end{aligned} \tag{16}$$

Exercise: Compute the density matrix of the first qubit if the total state is

$$|\Psi_2\rangle = \sqrt{\epsilon}|00\rangle + \sqrt{1-\epsilon}|11\rangle$$

3.2 The density matrix as a statistical mixture of pure states

We can manipulate the density matrix defined as a partial trace to express it as a statistical mixture of pure states. For this, all we have to do is a Schmidt decomposition (SVD) of $|\Psi_{\text{tot}}\rangle$:

$$|\Psi_{\text{tot}}\rangle = \sum_i s_i |\psi_i\rangle \otimes |\psi_{\text{env}}(i)\rangle.$$

This is shown in Fig. 8 (a). Plugging this decomposition into the partial trace definition of ρ , we arrive at

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|, \tag{17}$$

with $p_i = s_i^2$, where s_i are the Schmidt values obtained by performing a Schmidt decomposition of $|\Psi_{\text{tot}}\rangle$. This is shown in Fig. 8 (c). Crucially, we used the fact that the $|\psi_{\text{env}}(i)\rangle$ states form an orthonormal basis.

This form is known as the spectral decomposition of ρ . It is not unique (there exists other families $\{p'_i, \phi'_i\}$ such that $\rho = \sum_i p'_i |\phi'_i\rangle \langle \phi'_i|$ with $\sum_i p'_i = 1$. Note that $\{|\phi'_i\rangle\}$ need not be orthonormal). Using the unit trace property (Eq. (16)), we see that

$$\sum_i p_i = 1.$$

Therefore Eq. (17) can be interpreted as follows: the qubits are in a statistical mixture of pure states $\{|\psi_i\rangle\}$ with the associated probabilities $\{p_i\}$. In other words, we can say that the system is in state $|\psi_i\rangle$ with probability p_i . If it so happens that the total wavefunction factorizes, $|\Psi_{\text{tot}}\rangle = |\Psi_{\text{qubits}}\rangle \otimes |\Psi_{\text{environment}}\rangle$, then, using (15), we find that

$$\rho = |\Psi_{\text{qubits}}\rangle \langle \Psi_{\text{qubits}}|, \tag{18}$$

namely, if there is no entanglement between the qubits and the environment, the system is, with absolute certainty, in state $|\Psi_{\text{qubits}}\rangle$. *One thus sees that entanglement with the environment creates uncertainty in the knowledge of the state of the system.*

Mathematically, this can be measured by looking at the rank of the density matrix: a rank one means a “pure state” (as in Eq. (18)), a larger rank means a “mixed state” (Eq (17)).

Besides the rank, a measure of the mixed character of the density matrix is the so-called purity,

$$P = \text{Tr}(\rho^2). \quad (19)$$

For a pure state,

$$P = 1.$$

3.3 Single-qubit case

In the single-qubit case, the density matrix is a unit-trace 2×2 matrix so its decomposition on the Pauli basis is of the form

$$\rho = \frac{1}{2}I + \frac{1}{2}(r_x X + r_y Y + r_z Z) = \frac{I}{2} + \frac{1}{2}\mathbf{r} \cdot \boldsymbol{\sigma} \quad (20)$$

with $\boldsymbol{\sigma}$ the vector of Pauli matrices and $\mathbf{r} = (r_x, r_y, r_z)$

Let us compute

$$P = \text{Tr}(\rho^2) = \frac{1 + |\mathbf{r}|^2}{2},$$

thus $P = 1$ iff $|\mathbf{r}| = 1$ i.e $\mathbf{r}$ lives on the surface of a sphere. Thus pure states ($P = 1$) live on the surface of the Bloch sphere. For mixed states, $P < 1$ and thus $|\mathbf{r}| < 1$: mixed states live *inside* the Bloch sphere. Note also that in the worst case, $|\mathbf{r}| = 0$, corresponding to $P = 1/2$ and the state $\rho = I/2$: it is known as the maximally mixed state. This is the state where we know the least information about the system: it is in state $|0\rangle$ or in state $|1\rangle$ (as opposed to a pure state $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ where the system is in state $|0\rangle$ and $|1\rangle$).

3.4 Expressing the von Neumann entanglement entropy as a function of the subsystem density matrix

We are now in a good position to express the von Neumann entanglement entropy as a function of the subsystem density matrix. Recall that given a closed system composed of two subsystems A and B , the von Neumann entanglement entropy is expressed as a function of the Schmidt values of the Schmidt decomposition of the total wavefunction $|\Psi\rangle$ of the system (Eq. (10)).

Let us show that, equivalently, if we define, following Eq. (15),

$$\rho_A = \text{Tr}_B(|\Psi\rangle\langle\Psi|),$$

then we have:

$$S_{AB} = -\text{Tr}(\rho_A \log_2 \rho_A). \quad (21)$$

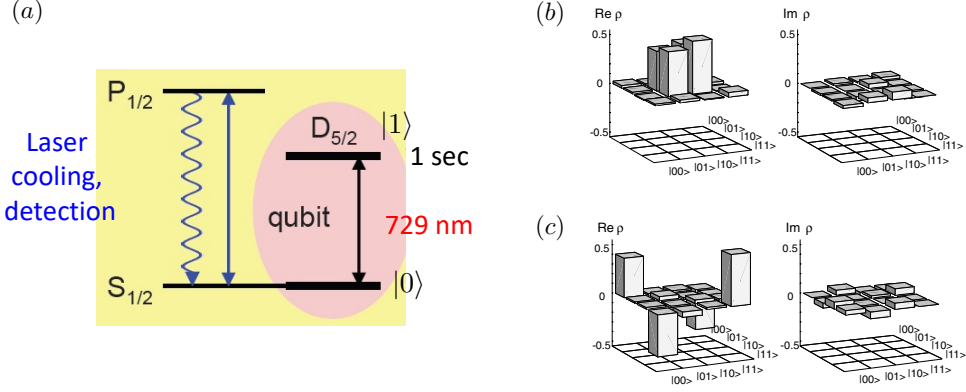


Figure 9: (a) State tomography on two Ca^+ ions encoding two qubit states $|0\rangle$ and $|1\rangle$ separated by a transition at 729 nm. (b) Results for the state: $(|01\rangle + |10\rangle)/\sqrt{2}$. (c) State: $(|00\rangle - |11\rangle)/\sqrt{2}$. From [Roos *et al.* \(2004\)](#).

Proof: we can switch from the partial trace definition of ρ_A to its spectral form, $\rho_A = \sum_i s_i^2 |\psi_i\rangle\langle\psi_i|$ where $|\psi_i\rangle$ forms an orthonormal basis of subsystem A . So the s_i^2 are also the eigenvalues of ρ_A , and therefore

$$S_{AB} = - \sum_i s_i^2 \log_2(s_i^2).$$

QED.

A practical consequence of the density-matrix based definition of S_{AB} , Eq. (21), is that one can in principle measure S_{AB} by measuring the density matrix ρ_A of subsystem A . This can be achieved via quantum state tomography.

3.5 Quantum state tomography

Performing the tomography of a quantum state means measuring its density matrix.

In the single-qubit case, for instance, three parameters, r_x , r_y and r_z are required to fully characterize the density matrix, as is clear from (20). In practice, these numbers can be recovered by measuring

$$\langle\sigma_\alpha\rangle = \text{Tr}(\rho\sigma_\alpha) = r_\alpha$$

for $\alpha = x, y, z$. (Check it!)

A two-qubit example is shown in Fig. 9.

For more qubits, however, it quickly gets very expensive since the number of degrees of freedom in ρ_A is 4^{n_A} (if n_A is the number of qubits in subsystem A ; this does not take into account the unit trace property of the density matrix).

More practical metrics for entropy have thus been introduced. This will be explained in Lecture 6.

Problem set

A Entanglement of random quantum states

This is a Python exercise.

1. Create a random quantum state using the `numpy.random` library.
2. Perform a SVD on the matrix corresponding to a reshaping of its amplitudes.
3. Plot the singular values. What do you notice? (Compare to Fig. 6)
4. Compute the von Neumann entropy using the singular values. Repeat the steps for increasing system size. What do you observe?
5. You should have obtained quite good agreement with a maximal entropy. What does the fact that it works for a single random quantum state (as opposed as when averaging over many samples) signal?

B Manipulating the density matrix

1. Using the partial trace definition of ρ , prove that

$$\langle O \rangle = \text{Tr} [\rho \hat{O}] \quad (22)$$

Proof: let us introduce a basis $|\psi_\alpha\rangle$ of the qubit space. Then

$$\begin{aligned} \text{Tr} [\rho \hat{O}] &= \sum_{\alpha} \langle \psi_{\alpha} | \sum_k \langle \varphi_k | \Psi_{\text{tot}} \rangle \langle \Psi_{\text{tot}} | \varphi_k \rangle \hat{O} | \psi_{\alpha} \rangle \\ &= \sum_{\alpha} \sum_k \langle \psi_{\alpha} | \langle \varphi_k | | \Psi_{\text{tot}} \rangle \langle \Psi_{\text{tot}} | \hat{O} | \psi_{\alpha} \rangle | \varphi_k \rangle \\ &= \text{Tr} [| \Psi_{\text{tot}} \rangle \langle \Psi_{\text{tot}} | \hat{O}] \\ &= \langle \Psi_{\text{tot}} | \hat{O} \otimes I | \Psi_{\text{tot}} \rangle \end{aligned}$$

Graphically, see Fig. 8 (b)

2. Derive the spectral form (24) without resorting to Penrose diagrams. Proof: If we decompose $|\Psi_{\text{tot}}\rangle = \sum_{\alpha k} a_{\alpha k} |\chi_{\alpha}\rangle |\varphi_k\rangle$, then Eq. (15) gives

$$\begin{aligned} \rho &= \sum_k \langle \varphi_k | \sum_{\alpha_1 k_1} a_{\alpha_1 k_1} |\chi_{\alpha_1}\rangle |\varphi_{k_1}\rangle \sum_{\alpha_2 k_2} a_{\alpha_2 k_2}^* \langle \chi_{\alpha_2} | \langle \varphi_{k_2} | \varphi_k \rangle \\ &= \sum_{\alpha_1} \sum_{\alpha_2} \left\{ \sum_k a_{\alpha_1 k} a_{\alpha_2 k}^* \right\} |\chi_{\alpha_1}\rangle \langle \chi_{\alpha_2}| \end{aligned}$$

In other words ρ is a matrix with coefficients

$$\rho_{\alpha_1 \alpha_2} = \sum_k a_{\alpha_1 k} a_{\alpha_2 k}^*,$$

in the $|\chi_\alpha\rangle$ basis. If we call A the $N \times M$ matrix with coefficients $a_{\alpha k}$ (where N (resp. M) is the size of the qubit (resp. environment) Hilbert space), then the above equation reads

$$\rho = AA^\dagger, \quad (23)$$

which makes it obvious that ρ is a positive semi-definite matrix. As such, we can diagonalize it as

$$\rho_{\alpha\alpha'} = \sum_i U_{\alpha i} p_i U_{i\alpha'}^\dagger$$

with $p_i \geq 0$. Let us now define the kets $|\psi_i\rangle$ such that $|\psi_i\rangle_\alpha = U_{\alpha i}$. Then,

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|. \quad (24)$$

3. Show that for any mixed state, the purity P obeys: $P \leq 1$. **Proof:**

$$\begin{aligned} \text{Tr} \rho^2 &= \text{Tr} \left(\sum_i p_i |\phi_i\rangle \langle \phi_i| \sum_j p_j |\phi_j\rangle \langle \phi_j| \right) \\ &= \sum_{ij} p_i p_j \text{Tr}(|\phi_i\rangle \langle \phi_i| |\phi_j\rangle \langle \phi_j|) \\ &= \sum_i p_i^2 \leq 1, \end{aligned}$$

where we have used the fact that $|\phi_i\rangle$ is orthonormal.

References

- Greenberger, Daniel M., Horne, Michael A., & Zeilinger, Anton. 1989. Going Beyond Bell's Theorem.
- Hastings, M B. 2007. An Area Law for One-Dimensional Quantum Systems. *Journal of Statistical Mechanics: Theory and Experiment*, **2007**(08), P08024–P08024.
- Roos, C. F., Lancaster, G. P. T., Riebe, M., Häffner, H., Hänsel, W., Gulde, S., Becher, C., Eschner, J., Schmidt-Kaler, F., & Blatt, R. 2004. Bell States of Atoms with Ultralong Lifetimes and Their Tomographic State Analysis. *Physical Review Letters*, **92**(22), 220402.
- White, Steven R. 1992. Density Matrix Formulation for Quantum Renormalization Groups. *Physical Review Letters*, **69**(19), 2863–2866.