Quantum information theory (20110401)

Lecturer: Jens Eisert

Chapter 2: Elements of quantum (information) theory

Contents

2 Elements of quantum (information) theory **5**
2.1 Quantum states and observables 5
2.1.1 Pure quantum states of qubits and qudits 5
2.1.2 Mixed quantum states of qubits 7
2.1.3 Quantum state spaces as convex sets 9
2.1.4 Observables and expectation values 10
2.1.5 Diagonalization and traces 12
2.2 Measurement postulate 12
2.3 Unitary time evolution 13
2.3.1 Schrödinger dynamics 13
2.3.2 Unitary operations 14
2.4 Composite quantum systems 14
2.4.1 Tensor products 14
2.4.2 Qubit registers 16
2.4.3 Partial traces 17
2.4.4 Schmidt decomposition for bi-partite pure quantum states . . 17

Chapter 2

Elements of quantum (information) theory

This is a course on quantum information theory, but it cannot hurt to brush up the basics of quantum theory, given that we will make heavy use of it. In this description, qubits and Pauli operators will feature. Therefore, in this chapter, we will be concerned with quantum mechanics as a physical theory. Every physical theory is supposed to make predictions on future measurement outcomes when performing experiments with a well-defined physical system that is initially prepared in the same way. The predictions of quantum mechanics are of a statistical nature: The theory is utterly silent about specific measurement outcomes. It will rather provide probabilities for obtaining certain outcomes. Conversely, to obtain evidence into the correctness of a prediction, one needs to perform many experiments under identical conditions. Then, by investigating relative frequencies of measurement outcomes, one can estimate probabilities. At the heart of the formalism are notions of expectation values. This is no shortcoming of the theory: This intrinsic randomness is actually a deep structure element of quantum mechanics that is there to stay: Bell's theorem shows that there cannot be an underlying classical statistical picture that can be held responsible to explain the randomness of quantum mechanics. That is to say, we have to ask ourselves how to capture *states* – the collection of information summarizing all information required to make future predictions – how *observables* – the quantities that can be measured. Also, we will think about how systems evolve in time. We will also see how composite quantum systems can be captured – very important in quantum information theory – and learn about the Schmidt decomposition.

2.1 Quantum states and observables

2.1.1 Pure quantum states of qubits and qudits

Classical bits can take the values 0 and 1 only. This is the commonly used basic unit of information, reflecting an on and off state of a basic cell. The analog of the

classical bit is the simplest quantum system, the *quantum bit* or in short *qubit*. This is the heart of the matter why quantum systems are more powerful than classical systems when it comes to applications in information processing. One can still identify 0 and 1 with basis vectors $|0\rangle$ and $|1\rangle$ of a complex vector space $\mathbb{C}^2$, but it state space is considerably bigger than that of classical bits. Quantum systems are associated with a *Hilbert space*, a complex Hilbert space is a vector space equipped with a scalar product that is complete with respect to the norm induced by the its scalar product. For a qubit, this vector space is $\mathbb{C}^2$.

State vectors: Pure quantum states are described by normalized state vectors $|\psi\rangle \in \mathcal{H}$ from a complex Hilbert space.

Such state vectors are general superpositions

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{2.1}$$

of basis vectors $|0\rangle$ and $|1\rangle$, with complex α, β, normalized as

$$|\alpha|^2 + |\beta|^2 = 1. \tag{2.2}$$

This example already shows that qubits cannot only be pointing up or down in quantum mechanics. They can be in an arbitrary superposition of pointing up or down. The associated Hilbert space is simply $\mathcal{H} = \mathbb{C}^2$. The scalar product between two state vectors is written as $\langle\psi|\phi\rangle$. Normalization means that the standard vector norm takes a unit value, which in turn is equivalent with

$$\langle\psi|\psi\rangle = 1. \tag{2.3}$$

The vector $\langle\psi| \in \mathcal{H}^*$ is a dual vector. Jokingly referring to the term bracket, one also calls dual vectors "bras" and vectors "kets". Matrix elements of operators A take the form $\langle\psi|A|\phi\rangle$. For every Hilbert space of a d-dimensional quantum system, referred to as *qudit* in the context of quantum information, one can pick a basis

$$\mathcal{B} = \{|0\rangle, \ldots, |d-1\rangle\}. \tag{2.4}$$

In this basis, every state vector can be expressed as

$$|\psi\rangle = \sum_{j=0}^{d-1} c_j |j\rangle. \tag{2.5}$$

The complex numbers $c_0, \ldots, c_{d-1}$ are called *coefficients*. The basis is normalized and complete, which means that

$$\langle j|k\rangle = \delta_{j,k}, \tag{2.6}$$

$$\sum_{j=0}^{d-1} |j\rangle\langle j| = \mathbb{1}. \tag{2.7}$$

All this applies to so-called finite-dimensional quantum systems, where d is an integer. We will see that qudits with a prime dimension d take a particularly important role. This has to do with the fact that $\mathbb{Z}_d$ is a finite field.

2.1.2 Mixed quantum states of qubits

As we know, pure states are not sufficient to capture all possible preparations. This is also true classically: More generally, we should consider the situation in which the bit takes the values 0 and 1 only with some probabilities p_0 and p_1, respectively. More formally put, the state space of a classical bit is the straight line segment, reflecting a "mixture" or a convex combination 0 and 1. If the probability of having 0 is p_0 and that of having one 1 is p_1, then the state of the system is given by a vector $(p_0, p_1) \in \mathbb{R}^2$ with

$$p_0, p_1 \geq 0 \tag{2.8}$$

normalized as

$$p_0 + p_1 = 1. \tag{2.9}$$

This may be a bit of an overloaded way of putting it: But this is a convex set, a simplex in fact, and $(1, 0)$ and $(0, 1)$ – corresponding to 0 and 1 – are the extreme points of this set. Probabilistic mixtures take values in the interior of the set.

Similarly, pure quantum states are not all there is in quantum mechanics. More generally, one needs to consider mixed states. A most paradigmatic situation is one in which one prepares $|0\rangle$ with some probability p_0 and $|1\rangle$ with probability $p_1 = 1 - p_0$. There is no state vector that reflects this situation. For this, we need to resort to *density operators*. The state space of a qubit is no longer a straight line segment, but can be represented as a ball, the *Bloch ball*. It generalizes probability distributions to matrices

$$\rho = \begin{bmatrix} p_0 & c \\ c^* & p_1 \end{bmatrix} \in \mathbb{C}^{2\times 2}. \tag{2.10}$$

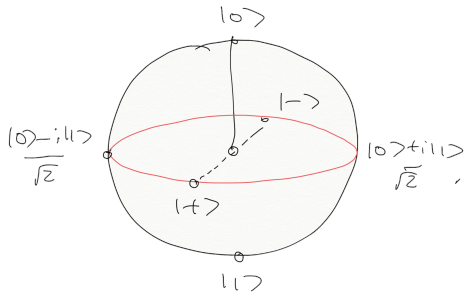
Eq. (2.8) is being replaced by the constraint that ρ is positive semi-definite,

$$\rho \geq 0, \tag{2.11}$$

that it is normalized as

$$\mathrm{tr}(\rho) = 1. \tag{2.12}$$

Such a matrix ρ is called *density matrix* or simply the *quantum state* of the qubit. Since this density matrix is obviously Hermitian, its main diagonal elements are clearly real, and they are positive by virtue of Eq. (2.11). In fact, due to Eq. (2.12), they can be identified with a classical probability distribution (p_0, p_1). In fact, diagonal density operators can be identified with finite probability distributions.



But there is more to a quantum state: There is now an off-diagonal element $c \in \mathbb{C}$ of ρ. This may be innocent looking, but makes a big difference. One can no longer interpret a quantum state as a classical alternative. It is not in a probabilistic mixture of 0 or 1. In fact, the off-diagonal blocks signify a superposition, the qubit can be in "0 and 1 at the same time". It is common for quantum systems to be in such superpositions, even if our everyday intuition may find this alien or strange. The extreme points are precisely the pure states that we know from elementary quantum mechanics. Again, these extreme points can be written as complex vectors

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{2.13}$$

with $\alpha, \beta \in \mathbb{C}$ satisfying $|\alpha|^2 + |\beta|^2 = 1$. The respective rank-1 projections onto $|0\rangle$ and $|1\rangle$ are then given by $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. These basis vectors are isomorphic to density operators as

$$|0\rangle \approx |0\rangle\langle 0| \approx \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \; |1\rangle \approx |1\rangle\langle 1| \approx \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.14}$$

It is both common to use the vector notation for pure states (i.e., extreme points of the set) as well as density matrices. It will depend on the context what is more natural to use. So indeed, a qubit has a larger state space than a simple bit, reflecting the superposition principle that is not present classically in the same fashion. This already points to the direction that we can use the qubit for the encoding of information in a different way than classically, even though there are some subtleties involved.

This is a good moment to discuss a number of examples. Let us go back to our initial situation discussed at the beginning of the chapter, of preparing $|0\rangle$ or $|1\rangle$ with equal probability. We can now easily associate this with a density operator

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|. \tag{2.15}$$

We can write this in matrix form – remember that operators and their matrix representation are identified with each other throughout the script

$$\rho = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{2.16}$$

We have that

$$\mathrm{tr}(\rho^2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1. \tag{2.17}$$

This in fact the minimum value $\mathrm{tr}(\rho^2)$ can take for a system with $\mathcal{H} = \mathbb{C}^2$. The pure state $\rho = |0\rangle\langle 0|$ in turn is represented as

$$\rho = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.18}$$

obviously satisfying $\mathrm{tr}(\rho) = 1$. Generally, if we have probabilities p_0 and p_1 to prepare $|0\rangle\rangle$ and $|1\rangle$, we have the density operator

$$\rho = \begin{bmatrix} p_0 & 0 \\ 0 & p_1 \end{bmatrix}. \tag{2.19}$$

But of course, we are not forced to take the standard basis. The situation of having prepared $|+\rangle$ and $|-\rangle$ with equal probabilities is captured as

$$\rho = \frac{1}{2}|+\rangle\langle+| + \frac{1}{2}|-\rangle\langle-|. \tag{2.20}$$

This is

$$\begin{aligned} \rho &= \frac{1}{4}\left((|0\rangle + |1\rangle)(\langle 0| + \langle 1|)\right) + \frac{1}{4}\left((|0\rangle - |1\rangle)(\langle 0| - \langle 1|)\right) \\ &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|, \end{aligned} \tag{2.21}$$

with matrix representation

$$\rho = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{2.22}$$

There is a cute observation that is important at this point: There are many different ways of preparing the same density operator

2.1.3 Quantum state spaces as convex sets

In the same way, general *density operators* or *quantum states* are positive semi-definite matrices over this complex vectors space $\mathbb{C}^d$ of dimension d.

General quantum states: A general quantum state of a d-dimensional quantum system is given by an operator ρ over the vector space $\mathbb{C}^d$ that is positive semi-definite and normalized as

$$\rho \geq 0, \;\; \mathrm{tr}(\rho) = 1. \tag{2.23}$$

The convex set of such operators is referred to as the *state space* $\mathcal{S}(\mathcal{H}) \subset \mathbb{C}^{d\times d}$.

It goes without saying that $\rho \geq 0$ already implies that

$$\rho = \rho^\dagger \tag{2.24}$$

is Hermitian. It means that its eigenvalues are real and non-negative. The familiar state vectors correspond to the pure states of this set.

Pure quantum states: The extreme points satisfy $\mathrm{tr}(\rho^2) = 1$ and correspond to vectors reflecting pure states, they can be written as

$$\rho = |\psi\rangle\langle\psi| \tag{2.25}$$

vectors $|\psi\rangle \in \mathcal{H}$ in that vector space, normalized as $\langle\psi|\psi\rangle = 1$.

The set $\mathcal{S}(\mathcal{H})$ is indeed a *convex set*: If $\rho_1 \in \mathcal{S}(\mathcal{H})$ and $\rho_2 \in \mathcal{S}(\mathcal{H})$, then the straight line segment

$$\lambda\rho_1 + (1-\lambda)\rho_2 \in \mathcal{S}(\mathcal{H}) \tag{2.26}$$

is again contained in state space. That is to say, the *mixing* (i.e., the convex combination) of two quantum states ρ_1 and ρ_2 gives again rise to a valid quantum state. The interpretation of mixing is basically the same as that of convex combinations of probability distributions. If we prepare ρ_i (pure or mixed) with probability p_i, for $i = 1, \ldots, n$, we obtain the mixed quantum state

$$\rho = \sum_{i=1}^{n} p_i \rho_i. \tag{2.27}$$

But again, only the physical quantum state ρ has any significance, and not the particular ensemble.

2.1.4 Observables and expectation values

Quantities that can be measured in experiments are called, unsurprisingly, *observables*. They are associated with Hermitian operators A, meaning that

$$A = A^\dagger. \tag{2.28}$$

Their eigenvalues (or rather spectral values, but let us not be too mathematically pedantic at this point) are possible outcomes of (idealized projective) measurements. The fact that observables are Hermitian implies the property that their eigenvalues (or spectral values) are real, which is a nice feature if one wants to interpret them as measurement outcomes.

Observables: Observable are Hermitian operations in a Hilbert space.

For pure states represented as state vectors $|\psi\rangle$, expectation values of such observables for systems prepared in pure states are given as

$$\langle A \rangle = \langle \psi | A | \psi \rangle. \tag{2.29}$$

From this, it is also obvious how to compute expectation values for general mixed states. Writing ρ as

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i| \tag{2.30}$$

and exploiting the linearity of the trace and the property that the trace is preserved under cyclic permutations, we find for the expectation value

$$\begin{aligned} \langle A \rangle &= \sum_{i=1}^{n} p_i \langle \psi | A | \psi \rangle & (2.31) \\ &= \sum_{i=1}^{n} p_i \mathrm{tr}(A|\psi_i\rangle\langle\psi_i) \\ &= \mathrm{tr}(A \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i) \\ &= \mathrm{tr}(A\rho). & (2.32) \end{aligned}$$

This insight is worth a box.

Expectation values: Expectation values of observables A of systems prepared in quantum states ρ are given by $\langle A\rangle = \mathrm{tr}(A\rho)$.

As we know, such expectation values make predictions about relative frequencies in experiments. An important family of observables that plays an important role in quantum infomation theory for good reason is that of the *Pauli matrices*.

Pauli matrices: The Pauli matrices are given by

$$X = |0\rangle\langle 1| + |1\rangle\langle 0| = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{2.33}$$

$$Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \tag{2.34}$$

$$Z = |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{2.35}$$

$$\mathbb{1} = |0\rangle\langle 0| + |1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{2.36}$$

The identity is commonly included in the collection of Pauli matrices. It is also now obvious how to compute their expectation values. The expectation value of Z of a system prepared in $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, for example, is simply given by

$$\langle\psi|Z|\psi\rangle = |\alpha|^2 - |\beta|^2. \tag{2.37}$$

The Pauli operators are not only observables, but also unitary operators.

It now also becomes clear what the significance is of quantum state spaces being no simplices: Since all expectation values of observables are computed as $\langle A\rangle = \mathrm{tr}(A\rho)$, we get exactly same same value for all observables in case of

$$\rho = \sum_{j=1}^{n} p_j|\psi_j\rangle\langle\psi_j| = \sum_{k=1}^{m} q_k|\phi_k\rangle\langle\phi_k|, \tag{2.38}$$

even if all of the probabilities $\{p_j\}$ and $\{q_k\}$ as well as all state vectors $\{|\psi_j\rangle\}$ and $\{|\phi_k\rangle\}$ are different. In fact, now even $n = m$ has to hold. What matters for all outcomes in all experiments is the density operator, not the mixed ensemble we have started with. Sometimes, people use notions of the kind, "the system *is* in some pure state vector $|\psi_j\rangle$, $j = 1, \ldots, n$, we simply do not know which one". Such reasoning is not quite precise and can be plain wrong, in which case it is referred to as *preferred ensemble fallacy*.

2.1.5 Diagonalization and traces

Observables can always be diagonalized. After all, unitary operators are precisely those that map one orthonormal basis onto another one. They are very important for all that is to come.[1]

Diagonalization: Every Hermitian $A \in \mathbb{C}^{d\times d}$ can be diagonalized in that there exists a unitary operator $U \in U(d)$ (satisfying $UU^\dagger = U^\dagger U = \mathbb{1}$) and a diagonal matrix D so that

$$A = UDU^\dagger. \tag{2.39}$$

That is to say, when expressed in the appropriate basis, every Hermitian operator takes a diagonal form in the matrix representation. General *operators* in Hilbert spaces can, needless to say, be expressed in this basis, as

$$A = \sum_{j,k=0}^{d-1} \langle j|A|k\rangle |j\rangle\langle k|. \tag{2.40}$$

Their *trace* is given by

$$\mathrm{tr}[A] = \sum_{j=0}^{d-1} \langle j|A|j\rangle. \tag{2.41}$$

The trace is independent of the choice of the basis, as a moment of thought reveals. The trace also has a further important property that we have actually already made use of above. For arbitrary operators A, B, and C, one has

$$\mathrm{tr}(ABC) = \mathrm{tr}(CAB), \tag{2.42}$$

for cyclic permutations of the operators.

2.2 Measurement postulate

The measurement postulate has to give an answer to the following questions: What are the outcomes of a measurement? What is the probability of obtaining this? What is the state immediately after the measurement? The measurement postulate settles these questions for so-called *von-Neumann measurements*. We will soon turn to a slightly more general picture that we need a lot in quantum information.

[1] It may be worth noting that the $d \times d$ unitaries form a group, the group $U(d)$.

Measurement postulate: Let A be an observable with spectral decomposition

$$A = \sum_k \lambda_k P_k, \tag{2.43}$$

where

$$P_k = \sum_{\mathrm{EV}\lambda_k} |\psi_k\rangle\langle\psi_k|. \tag{2.44}$$

That is, the $P_k = P_k^2$ are the projections onto the eigenspaces to eigenvalue λ_k. The possible measurement outcomes are λ_j. The probability of obtaining for obtaining the outcome related to λ_k is

$$p_k = \mathrm{tr}[\rho P_k]. \tag{2.45}$$

The state immediately after the measurement is

$$\rho_k' = \frac{P_k \rho P_k}{\mathrm{tr}[P_k \rho P_k]} = \frac{P_k \rho P_k}{\mathrm{tr}[\rho P_k]}. \tag{2.46}$$

2.3 Unitary time evolution

2.3.1 Schrödinger dynamics

When we prepare a quantum system in some state ρ, how does it evolve in time? The answer to this question is given by the Schrödinger equation. It is given in the form of a differential equation. For state vectors, we know the following expression capturing Schrödinger dynamics generated by some Hamiltonian H,

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H|\psi(t)\rangle, \tag{2.47}$$

from which we immediately find the following von-Neumann equation.

Von-Neumann equation: General quantum states evolve as

$$i\hbar \frac{\partial}{\partial t} \rho(t) = [H, \rho(t)]. \tag{2.48}$$

Making use of the unitary *time evolution operator*

$$U(t) = e^{-iHt/\hbar}, \tag{2.49}$$

valid for time-independent Hamiltonians, we can capture time evolution also as follows. Since it is unitary, it satisfies

$$U(t)U^\dagger(t) = U^\dagger(t)U(t) = \mathbb{1}. \tag{2.50}$$

Obviously, $U(0) = \mathbb{1}$.

Time evolution: The time evolution of a closed quantum systems from time t_1 to $t_2 > t_1$ is captured by a time evolution of state vectors as

$$|\psi(t_2)\rangle = U(t_2 - t_1)|\psi(t_1)\rangle. \tag{2.51}$$

General states, so density operators, evolve according to

$$\rho(t_2) = U(t_2 - t_1)\rho(t_1)U^\dagger(t_2 - t_1). \tag{2.52}$$

This type of time evolution is referred to as time evolution in the *Schrödinger* picture, in which observables are kept constant and quantum states evolve. It can also make sense to refer to a picture in which states following preparations are kept constant and observables evolve. This picture is referred to as *Heisenberg picture*. Of course, the predictions in both pictures are identical. One way of putting time evolution, therefore, is to say that states at different times are unitarily equivalent, and all the Schrödiger equation does is to rotate the frame.

2.3.2 Unitary operations

There is an important shift in mindset, however, that is commonly made in quantum information theory compared to that of common introductory courses: In the latter, one usually sees the Hamiltonian H as the fundamental object, and one aims for identifying how the system naturally evolves in time. We are still dealing with the same quantum mechanics, but one looks at the problem differently. One sees

$$\rho \mapsto U\rho U^\dagger \tag{2.53}$$

as a *unitary operation* of a quantum state ρ. Physically speaking, this is of course reflecting nothing but unitary time evolution: But one sees U as the central object, which could then have been generated from a Hamiltonian as $U = e^{iHt}$ for some time $t \geq 0$. One often abstracts from these Hamiltonians, however, and thinks of manipulating a given state with a unitary. We will later encounter quantum gates are particularly important instances of basic unitary primitives. Since the Pauli operators are unitary, the constitute important examples in this respect.

2.4 Composite quantum systems

2.4.1 Tensor products

How do we describe composite quantum systems in quantum theory? Clearly, the formalism must have an answer to that. We think of a particle having several degrees of freedom. Or we aim at describing several different particles at once. How do we capture this situation? Composition of degrees of freedom is incorporated by the tensor

products in quantum mechanics. Let us assume that we have one degree of freedom associated with a d_1-dimensional Hilbert space

$$\mathcal{H}_1 = \text{span}\{|0\rangle, \ldots, |d_1 - 1\rangle\}. \tag{2.54}$$

We then consider another, second degree of freedom, coming along with a d_2-dimensional Hilbert space

$$\mathcal{H}_2 = \text{span}\{|0\rangle, \ldots, |d_2 - 1\rangle\}. \tag{2.55}$$

These spaces could, for example, capture all superpositions of two qubit degrees of freedom of two particles described by quantum mechanics. The Hilbert space of the *joint system* is then given by the *tensor product*[2]

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{2.59}$$

It is spanned by the orthonormal basis vectors

$$\{|j\rangle \otimes |k\rangle : j = 0, \ldots, d_1 - 1; k = 0, \ldots, d_2 - 1\}. \tag{2.60}$$

Such basis elements of tensor products are sometimes also written as

$$\{|j, k\rangle : j = 0, \ldots, d_1 - 1; k = 0, \ldots, d_2 - 1\}. \tag{2.61}$$

This looks more complicated than it is: While an arbitrary superposition of a state vector from $\mathcal{H}_1$ can be written as

$$|\psi_1\rangle = \sum_{j=0}^{d_1-1} \alpha_j |j\rangle \tag{2.62}$$

and an arbitrary superposition of a state vector from $\mathcal{H}_2$ is

$$|\psi_2\rangle = \sum_{j=0}^{d_2-1} \beta_j |j\rangle, \tag{2.63}$$

an arbitrary state vector taken from the composite Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is given by

$$|\psi\rangle = \sum_{j=0}^{d_1-1} \sum_{k=0}^{d_2-1} \gamma_{j,k} |j\rangle \otimes |k\rangle, \tag{2.64}$$

as a linear combination of all new basis vectors, with all $\gamma_{j,k} \in \mathbb{C}$. If you think at this point that it may be confusing that such general state vectors contain ones that are no

[2]Basic linear algebraic properties of the tensor product are taken for granted in this course. E.g., tensor products satisfy

$$|\psi\rangle \otimes |\omega\rangle + |\phi\rangle \otimes |\omega\rangle = (|\psi\rangle + |\phi\rangle) \otimes |\omega\rangle, \tag{2.56}$$

$$|\omega\rangle \otimes |\psi\rangle + |\omega\rangle \otimes |\phi\rangle = |\omega\rangle \otimes (|\psi\rangle + |\phi\rangle), \tag{2.57}$$

$$\alpha|\psi\rangle \otimes |\phi\rangle = (\alpha|\psi\rangle) \otimes |\phi\rangle = |\psi\rangle \otimes (\alpha|\phi\rangle) \tag{2.58}$$

for $\alpha \in \mathbb{C}$ and $|\psi\rangle$, $\phi\rangle$, $|\omega\rangle$ being state vector of their respective vector spaces.

longer a product between the respective Hilbert spaces: Indeed, it is, and we will come to the profound implications of this later. Again:

Hilbert spaces of composite quantum systems: The Hilbert space of the composite quantum systems the parts being associated with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ is given by the tensor product

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{2.65}$$

2.4.2 Qubit registers

A particularly important situation, unsurprisingly, is the one where we have several or many qubits at hand: They could be quantum registers in a protocol in quantum metrology, quantum key distribution. Or they could serve as computational registers of a quantum computer. This situation is captured by the n-fold tensor product.

Qubit registers: The Hilbert space of n qubits is given by $(\mathbb{C}^2)^{\otimes n} = \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$.

That is to say, arbitrary state vectors of n qubits can be written as

$$\begin{aligned} |\psi\rangle &= \alpha_{0,\ldots,0,0}|0\rangle \otimes \cdots \otimes |0\rangle \otimes |0\rangle + \alpha_{0,\ldots,0,1}|0\rangle \otimes \cdots \otimes |0\rangle \otimes |1\rangle \\ &+ \ldots \alpha_{1,\ldots,1,1}|1\rangle \otimes \cdots \otimes |1\rangle \otimes |1\rangle. \end{aligned} \tag{2.66}$$

A basis of $\mathcal{H}$ is

$$\mathcal{B} := \{|i_1\rangle \otimes \cdots \otimes |i_{n-1}\rangle \otimes |i_n\rangle,\ i_1, \ldots, i_n \in \{0,1\}\}. \tag{2.67}$$

Since the many tensor products can be clumsy, one often writes $|0,\ldots,0,0\rangle$ instead of $|0\rangle \otimes \cdots \otimes |0\rangle \otimes |0\rangle$. Again, an *arbitrary superposition* of basis vectors as in Eq. (2.66) is a legitimate state vector corresponding to a pure state. This reflects the situation that a collection of qubits can – in a sense – be in "all classical alternatives at once". This idea is also at the heart of quantum computing, in that a register is simultaneously manipulated in a superposition state reflecting several inputs "at once". The precise functioning is subtle and more complicated than that, but this statement already creates the right mental image to see what this is about. We come back to this at the end of this section. Similarly, an arbitrary linear operator can be decomposed as

$$O = \sum_{j,k} c_{j,k} A_j \otimes B_k, \tag{2.68}$$

with operators $\{A_j\}$ and $\{B_k\}$ on $\mathcal{H}_1$ and $\mathcal{B}_1$, respectively. We will frequently encounter, e.g., tensor products of Pauli operators, such as

$$X \otimes X,\ Y \otimes Z, \tag{2.69}$$

over $(\mathbb{C}^2)^{\otimes 2}$ for two qubits. We will see that the tensor products of Pauli operators equipped with the right pre-factors form a group, the *Pauli group*, which is important in many aspects of quantum information theory, in particular for quantum error correction in quantum computing.

2.4.3 Partial traces

We have encountered the trace, but a concept that is also important in quantum information theory is the partial trace. This is particular important for the concept of a *reduced quantum state*: It is the quantum mechanical analog of a marginal distribution. It is the quantum state that one associates to a system if one decides to ignore another part of the quantum system. Surely, one has to find a consistent assignment. The reduced state does that.

Formally speaking, in a composite quantum systems with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ with bases $\mathcal{B}_1$ and $\mathcal{B}_2 = \{|0\rangle, \ldots, |d-1\rangle\}$, and an operator O on $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ has the partial trace "to 1" or "tracing out 2" given by

$$\mathrm{tr}_2(A) = \sum_{j=0}^{d-1} (\mathbb{1} \otimes \langle j|) O (\mathbb{1} \otimes |j\rangle). \tag{2.70}$$

Is is just that: A partial trace over only one tensor factor. Note that while the definition depends on a specific choice of a basis in $\mathcal{B}_2$, it is easy to see that the partial trace is invariant under the specific basis chosen. Again, particularly important are reduced states

$$\rho_1 := \mathrm{tr}_2(\rho) \tag{2.71}$$

of quantum states ρ of a composite quantum system. It is also clear that any expectation value of any observable A acting only in the first tensor factor 1 can be computed from the reduced state only: A moment of thought reveals that

$$\mathrm{tr}(A\rho_1) = \mathrm{tr}((A \otimes \mathbb{1})(\rho)). \tag{2.72}$$

In the light of this observation, the intuition of ignoring a part of a composite quantum system is most manifest. Let us consider an example. The reduced state of a pure state $|\psi\rangle\langle\psi|$ with state vector

$$|\psi\rangle = (|0,0\rangle + |1,1\rangle)/\sqrt{2} \tag{2.73}$$

of the first tensor factor is given by

$$\begin{aligned} \mathrm{tr}_2|\psi\rangle\langle\psi| &= \frac{1}{2}(\mathbb{1} \otimes \langle 0|)(|0,0\rangle + |1,1\rangle)(\langle 0,0| + \langle 1,1|)(\mathbb{1} \otimes |0\rangle) \quad (2.74) \\ &+ \frac{1}{2}(\mathbb{1} \otimes \langle 1|)(|0,0\rangle + |1,1\rangle)(\langle 0,0| + \langle 1,1|)(\mathbb{1} \otimes |1\rangle) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|). \quad (2.75) \end{aligned}$$

Interestingly, this is no longer a pure state: In fact, it is maximally mixed, the state that is in the centre of the Bloch ball. This is a purely quantum phenomenon, reflecting entanglement: Reduced states of pure states can be mixed.

2.4.4 Schmidt decomposition for bi-partite pure quantum states

We end this chapter with a particularly useful insight into pure states of systems composed of two parts. Such systems are called *bi-partite quantum systems*. Since they are

often seen as reflecting distributed quantum systems that can be locally manipulated by two experimentalists, Alice and Bob, say, the subsystems are commonly labeled A and B instead of 1 and 2. This situation is indeed highly instructive and important and we look at this in great detail. We consider bi-partite systems of dimension d each. A product basis of its Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is given by

$$\mathcal{B} = \{|j,k\rangle, j,k = 0,\ldots,d-1\}. \tag{2.76}$$

A pure state vector can be written as

$$|\psi\rangle = \sum_{j,k=0}^{d-1} c_{j,k}|j,k\rangle, \tag{2.77}$$

as we know. This set of state vectors includes *product states* that can be written as

$$|\psi\rangle = |\phi\rangle \otimes |\omega\rangle. \tag{2.78}$$

E.g., a product of basis vectors $|\psi\rangle = |0,0\rangle$ would be such a product state vector. Such states do not feature any correlations once one performs measurement on the respective parts: The probability distribution generated in this fashion will feature no correlations. Other vectors that are not products will feature correlations and are called *entangled*. We will hear a lot more about entanglement. These are quantum correlations that in a precise sense have no classical analog. It also means that probability distributions arising from measurements will be correlated. For our purposes important is the fact that there exists a basis such that the state vector takes a particularly simple form.

Schmidt decomposition: For every state vector $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ there exists bases $\mathcal{C}_A = \{|a_j\rangle\}$ and $\mathcal{C}_B = \{|b_j\rangle\}$ of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, such that

$$|\psi\rangle = \sum_{j=0}^{d-1} \sqrt{\lambda_j}|a_j, b_j\rangle, \tag{2.79}$$

with $\lambda_j \geq 0$ for all $j = 0,\ldots,d-1$ and

$$|\psi\rangle = (U \otimes V) \sum_{j,k=0}^{d-1} c_{j,k}|j,k\rangle, \ \sum_{j=0}^{d-1} \lambda_j = 1. \tag{2.80}$$

The interesting feature of this form is that the double sum is now replaced by a single sum: It is a normal form every pure state vector can be brought into. Note also that since going from one local basis to another is reflected by a unitary basis change by virtue of unitaries U and V acting on the respective tensor factors, the above statement can also be equivalently phrased in terms of the existence of such $U, V \in U(d)$ such that

$$(U \otimes V)|\psi\rangle = \sum_{j=0}^{d-1} \sqrt{\lambda_j}|j,j\rangle \tag{2.81}$$

Since the Schmidt form is a normal form, all the "non-local content" can be read off from it. For example, a maximally entangled pure state is one for which the reduced state to A is a *maximally mixed state*,

$$\mathrm{tr}_2(|\psi\rangle\langle\psi|) = \frac{1}{d}\mathbb{1}. \tag{2.82}$$

This can again be simply read off the Schmidt decomposition.

Schmidt form of maximally entangled state vectors: A maximally entangled state vector is one for which

$$\lambda_j = \frac{1}{d} \tag{2.83}$$

for all $j = 0, \ldots, d-1$, a product state vector satisfies $\lambda_0 = 1$ and $\lambda_1, \ldots, \lambda_{d-1} = 0$.

The proof of the Schmidt decomposition is a simple application of the singular value decomposition. Let us first state it in general terms.

Singular value decomposition: For any $A \in \mathbb{C}^{d_1 \times d_2}$, there exist $U \in U(d_1)$ and $V \in U(d_2)$ such that

$$UAV = D \tag{2.84}$$

where D is a non-negative real diagonal matrix.

The entries of D are called *singular values*. For positive semi-definite Hermitian matrices, these are nothing but the eigenvalues. For general Hermitian matrices, they are the absolute values of the eigenvalues. But singular values make sense for arbitrary matrices, and they are hugely important, to say the least. We will come back to them later. We are now prepared to prove the Schmidt decomposition. Let $C \in \mathbb{C}^{d \times d}$ be the matrix composed of the coefficients $\{c_{j,k}\}$ in Eq. (2.80). By virtue of the singular value decomposition, there exist $U, V \in U(d)$ such that

$$C = UDV, \tag{2.85}$$

where

$$D = \mathrm{diag}(\sqrt{\lambda_0}, \ldots, \sqrt{\lambda_{d-1}}). \tag{2.86}$$

Note the reversed roles of U and $U^\dagger$ as well as of V and $V^\dagger$, but this simplifies our

notation. We therefore find

$$\begin{aligned}
\sum_{j,k=0}^{d-1} c_{j,k}|j,k\rangle &= \sum_{j,k,l,m} U_{j,l} D_{l,m} V_{m,k}|j,k\rangle \qquad (2.87)\\
&= \sum_{j,k,l} U_{j,l}\sqrt{\lambda_l} V_{l,k}|j,k\rangle \\
&= \sum_{l=0}^{d-1} \sqrt{\lambda_l}\left(\sum_{j=0}^{d-1} U_{j,l}|j\rangle\right) \otimes \left(\sum_{k=0}^{d-1} V_{l,k}|k\rangle\right) \\
&= \sum_{l=0}^{d-1} \sqrt{\lambda_l}|a_l\rangle \otimes |b_l\rangle. \qquad (2.88)
\end{aligned}$$

Here, we have used that if $U \in U(d)$, then also the Hermitian conjugates $U^\dagger \in U(d)$ and the transpose $U^T = U(d)$ are unitary. Hence, it is easy to see that the $\{|a_l\rangle\}$ and $|b_l\rangle$ indeed constitute bases.