Quantum Information and Quantum Noise

Gabriel T. Landi
University of São Paulo

June 12, 2019

Contents

Chapter 1

Review of Quantum Theory

Quantum mechanics is about **states**, **operations** and **measurement outcomes**. You have probably already learned all three concepts in the old-fashioned way. Here you will learn it the *cool* way.

States represent the configuration of your system at any given time. You have probably seen them in the form of kets, such as $|\psi\rangle$ and $|i\rangle$, or as wave-functions $\psi(x)$. We will soon learn that the most general state in quantum mechanics is given by a new kind of object, called a **density matrix**, ρ. It encompass all the information you may have about a physical system, with kets and wave-functions being simply particular cases. Density matrices are absolutely essential for any modern use of quantum theory, which is why this chapter was named after them.

Once states are defined, it is natural to talk also about operations and outcomes. By operation I mean *any action which acts on the state to produce a new state*. This can be given, for instance, in terms of quantum mechanical operators such as H, $\hat{p}$, $a^\dagger$, σ_z, etc. Understanding operations from an intuitive sense is the key to understanding the physics behind the mathematics. After you gain some intuition, by simply looking at a Hamiltonian you will already be able to draw a bunch of conclusions about your system, without having to do any calculations.

Operators also fall into different categories, depending on what they are designed to do. The two most important classes are Hermitian and Unitary operators. Hermitian operators always have real eigenvalues and are used to describe quantities that can be observed in the lab. Unitary operators, on the other hand, preserve probabilities for kets and are used to describe the evolution of closed quantum systems. The evolution of an open quantum system, on the other hand, is described by another type of process known as **Quantum Operations** or **Quantum channels**, where instead of operators we use *super-operators* (which, you have to admit, sounds cool).

Finally, we have measurements and **outcomes**. Measurements are also implemented by operators. For instance, the wave-function collapse idea is what we call a *projective measurements* and is implemented by a projection operator. In this chapter we will learn how to generalize this and construct **generalized measurements and POVMs**. Quantum measurements is perhaps one of the most conceptually challenging process in quantum theory. And the theory of POVMs will help shed considerable light it.

1.1 Hilbert spaces and kets

"To any physical system we can associated an abstract complex vector space with inner product, known as a **Hilbert space***, such that the state of the system at an given instant can be described by a vector in this space."* This is the first and most basic postulate of quantum mechanics. Following Dirac, we usually denote vectors in this space as $|\psi\rangle$, $|i\rangle$, etc. The quantity inside the $|\rangle$ is nothing but a *label* to specify something.

A Hilbert space can be both finite or infinite dimensional. The dimension d is defined by the number of linearly independent vectors we need to span the vector space. A set $\{|i\rangle\}$ of linearly independent vectors that spans the vector space is called a basis. With this basis any state may be expressed as

$$|\psi\rangle = \sum_{i=0}^{d-1} \psi_i |i\rangle, \tag{1.1}$$

where ψ_i can be arbitrary complex numbers.

A Hilbert space is also equipped with an **inner product**, $\langle\phi|\psi\rangle$, which converts pairs of vectors into complex numbers, according to the following rules:

1. If $|\psi\rangle = a|\alpha\rangle + b|\beta\rangle$ then $\langle\gamma|\psi\rangle = a\langle\gamma|\alpha\rangle + \langle\gamma|\beta\rangle$.
2. $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$.
3. $\langle\psi|\psi\rangle \geq 0$ and $\langle\psi|\psi\rangle = 0$ if and only if $|\psi\rangle = 0$.

A set of basis vectors $|i\rangle$ is called **orthonormal** when it satisfies

$$\langle i|j\rangle = \delta_{i,j}. \tag{1.2}$$

Exploring the 3 properties of the inner product, one may then show that given two states written in this basis, $|\psi\rangle = \sum_i \psi_i |i\rangle$ and $|\phi\rangle = \sum_i \phi_i |i\rangle$, the inner product becomes

$$\langle\psi|\phi\rangle = \sum_i \psi_i^* \phi_i. \tag{1.3}$$

We always work with orthonormal bases. And even though the basis set is never unique, the basis we are using is usually clear from the context. A general state such as (1.1) is then generally written as a column vector

$$|\psi\rangle = \begin{pmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \psi_{d-1} \end{pmatrix}. \tag{1.4}$$

The object $\langle\psi|$ appearing in the inner product, which is called a **bra**, may then be written as a row vector

$$\langle\psi| = \begin{pmatrix} \psi_0^* & \psi_1^* & \dots & \psi_{d-1}^* \end{pmatrix}. \tag{1.5}$$

The inner product formula (1.3) can now be clearly seen to be nothing but the multiplication of a row vector by a column vector. Notwithstanding, I am obligated to emphasize that when we write a state as in Eq. (1.4), we are making specific reference to a basis. If we were to use another basis, the coefficients would be different. The inner product, on the other hand, does not depend on the choice of basis. If you use a different basis, each term in the sum (1.3) will be different, but the total sum will be the same.

The vectors in the Hilbert space which represent physical states are also constructed to satisfy the **normalization** condition

$$\langle\psi|\psi\rangle = 1. \tag{1.6}$$

This, as we will see, is related to the probabilistic nature of quantum mechanics. It means that if two states differ only by a global phase $e^{i\theta}$, then they are physically equivalent. Mathematically, this means that the relevant space in question is not a vector space, but rather a **ray space**.

You may also be wondering about wave-functions. Wave-functions are nothing but the inner product of a ket with the position state $|x\rangle$:

$$\psi(x) = \langle x|\psi\rangle. \tag{1.7}$$

The ket $|x\rangle$ may at first seem weird because the label inside the ket varies continuously. But apart from that, you can use it just as a normal ket. Wave-functions are not very useful in quantum information. In fact, I don't think we will ever need them again in this course. So bye-bye $\psi(x)$.

1.2 Qubits and Bloch's sphere

The simplest quantum system is one whose Hilbert space has dimension $d = 2$, which is what we call a **qubit**. In this case we only need two states that are usually labeled as $|0\rangle$ and $|1\rangle$ and are often called the **computational basis**. It is *very* important to understand that when we refer to a qubit, we don't necessarily have to mention which physical system it represents. For instance, it may represent a spin 1/2 particle, an atom with only two levels or the polarization directions of a photon. But when we say qubit, we mean neither. We just mean the underlying mathematical structure. This is the difference between a mathematical theory and the physical implementations. Quantum information is *independent* of the physical implementation. That is not to say that implementation s are not important. *Of course* they are! Without experiments all this theory would be meaningless. But that being said, one of nicest things about quantum information is that it is constructed independent of the implementation.

An arbitrary state of a qubit may be written as

$$|\psi\rangle = a|0\rangle + b|1\rangle = \begin{pmatrix} a \\ b \end{pmatrix}, \tag{1.8}$$

where a and b are complex numbers which, according to Eq. (1.6), should satisfy

$$|a|^2 + |b|^2 = 1 \tag{1.9}$$

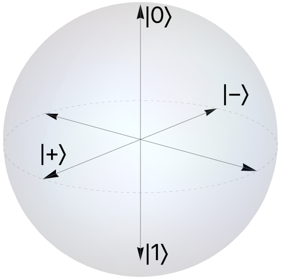

Figure 1.1: Example of Bloch's sphere which maps the general state of a qubit into a sphere of unit radius.

A convenient way to parametrize a and b in order to satisfy this normalization, is as

$$a = \cos(\theta/2), \qquad b = e^{i\phi} \sin(\theta/2), \tag{1.10}$$

where θ and ϕ are arbitrary real parameters. While this parametrization may not seem unique, it turns out that it is since any other choice will only differ by a global phase and hence will be physically equivalent. It also suffices to consider the parameters in the range $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$, as other values would just give the same state up to a global phase.

You can probably see a similarity here with the way we parametrize a sphere in terms of a polar and a azimutal angle. This is somewhat surprising since these are completely different things. A sphere is an object in $\mathbb{R}^3$, whereas in our case we have a vector in $\mathbb{C}^2$. But since our vector is constrained by the normalization (1.9), it is possible to map one representation into the other. That is the idea of **Bloch's sphere**, which is illustrated in Fig. 1.1. In this representation, the state $|0\rangle$ is the north pole, whereas $|1\rangle$ is the south pole. I also highlight in the figure two other states which appear often, called $|\pm\rangle$. They are defined as

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}. \tag{1.11}$$

In terms of the angles θ and ϕ in Eq. (1.10), this corresponds to $\theta = \pi/2$ and $\phi = 0, \pi$. Thus, these states lie in the *equator*, as show in Fig. 1.1.

A word of warning: Bloch's sphere is only used as a way to represent a complex vector as something real, so that we humans can visualize it. Be careful not to take this mapping too seriously. For instance, if you look blindly at Fig. 1.1 you will think $|0\rangle$ and $|1\rangle$ are parallel to each other, whereas in fact they are orthogonal, $\langle 0|1\rangle = 0$.

1.3 Outer product and completeness

The inner product gives us a recipe to obtain numbers starting from vectors. As we have seen, to do that, we simply multiply row vectors by column vectors. We could also think about the opposite operation of multiplying a column vector by a row vector. The result will be a matrix. For instance, if $|\psi\rangle = a|0\rangle + b|1\rangle$ and $|\phi\rangle = c|0\rangle + d|1\rangle$, then

$$|\psi\rangle\langle\phi| = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c^* & d^* \end{pmatrix} = \begin{pmatrix} ac^* & ad^* \\ bc^* & bd^* \end{pmatrix}. \tag{1.12}$$

This is the idea of the **outer product**. In linear algebra the resulting object is usually referred to as a rank-1 matrix.

Let us go back now to the decomposition of an arbitrar state in a basis, as in Eq. (1.1). Multiplying on the left by $\langle j|$ and using the orthogonality (1.2) we see that

$$\psi_i = \langle i|\psi\rangle. \tag{1.13}$$

Substituting this back into Eq. (1.1) then gives

$$|\psi\rangle = \sum_i |i\rangle\langle i|\psi\rangle.$$

This has the form $x = ax$, whose solution must be $a = 1$. Thus

$$\boxed{\sum_i |i\rangle\langle i| = 1 = \mathbb{I}} \tag{1.14}$$

This is the **completeness relation**. It is a direct consequence of the orthogonality of a basis set: all orthogonal bases satisfy this relation. In the right-hand side of Eq. (1.14) I wrote both the symbol $\mathbb{I}$, which stands for the identity matrix, and the number 1. A rigorous person would use $\mathbb{I}$. I'm sloppy, so I just write it as the number 1. I know it may seem strange to use the same symbol for a matrix and a number. But if you think about it, both satisfy *exactly* the same properties, so its not really necessary to distinguish them.

To make the idea clearer, consider first the basis $|0\rangle$ and $|1\rangle$. Then

$$|0\rangle\langle 0| + |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is the completeness relation, as expected since $|0\rangle, |1\rangle$ form an orthonormal basis. But we can also do this with other bases. For instance, the states (1.11) also form an orthogonal basis, as you may check. Hence, they must also satisfy completeness:

$$|+\rangle\langle +| + |-\rangle\langle -| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The completeness relation (1.14) has an important interpretation in terms of **projection onto orthogonal subspaces**. Given a Hilbert space, one may sub-divide it into several sub-spaces of different dimensions. The number of basis elements that you need to span each sub-space is called the rank of the sub-space. For instance, the space spanned by $|0\rangle$, $|1\rangle$ and $|2\rangle$ may be divided into a rank-1 sub-spaced spanned by the basis element $|0\rangle$ and a rank-2 sub-space spanned by $|1\rangle$ and $|2\rangle$. Or it may be divided into 3 rank-1 sub-spaces.

Each term in the sum in Eq. (1.14) may now be thought of as a projection onto a rank-1 sub-space. In fact, we define **rank-1 projectors**, as operators of the form

$$P_i = |i\rangle\langle i|. \tag{1.15}$$

They are called projection operators because if we apply them onto a general state of the form (1.1), they will only take the part of $|\psi\rangle$ that lives in the sub-space $|i\rangle$:

$$P_i|\psi\rangle = \psi_i|i\rangle.$$

They also satisfy

$$P_i^2 = P_i, \qquad P_i P_j = 0 \quad \text{if } i \neq j, \tag{1.16}$$

which are somewhat intuitive: if you project twice, you gain nothing new and if you project first on one sub-space and then on another, you get nothing since they are orthogonal.

We can construct projection operators of higher rank simply by combining rank-1 projectors. For instance, the operator $P_0 + P_{42}$ projects onto a sub-space spanned by the vectors $|0\rangle$ and $|42\rangle$. An operator which is a sum of r rank-1 projectors is called a rank-r projector. The completeness relation (1.14) may now also be interpreted as saying that if you project onto the full Hilbert space, it is the same as not doing anything.

1.4 Operators

The outer product is our first example of a linear operator. That is, an operator that acts linearly on vectors to produce other vectors:

$$A\bigg(\sum_i \psi_i|i\rangle\bigg) = \sum_i \psi_i A|i\rangle.$$

Such a linear operator is completely specified by knowing its action on all elements of a basis set. The reason is that, when A acts on an element $|j\rangle$ of the basis, the result will also be a vector and must therefore be a linear combination of the basis entries:

$$A|j\rangle = \sum_i A_{i,j}|i\rangle \tag{1.17}$$

The entries $A_{i,j}$ are called the matrix elements of the operator A in the basis $|i\rangle$. The quickest way to determine them is by taking the inner product of Eq. (1.17) with $\langle j|$, which gives

$$A_{i,j} = \langle i|A|j\rangle. \tag{1.18}$$

We can also use the completeness (1.14) twice to write

$$A = 1A1 = \sum_{i,j} |i\rangle\langle i|A|j\rangle\langle j| = \sum_{i,j} A_{i,j}|i\rangle\langle j|. \tag{1.19}$$

We therefore see that the matrix element $A_{i,j}$ is the coefficient multiplying the outer product $|i\rangle\langle j|$. Knowing the matrix form of each outer product then allows us to write A as a matrix. For instance,

$$A = \begin{pmatrix} A_{0,0} & A_{0,1} \\ A_{1,0} & A_{1,1} \end{pmatrix} \tag{1.20}$$

Once this link is made, the transition from abstract linear operators to matrices is simply a matter of convenience. For instance, when we have to multiply two linear operators A and B we simply need to multiply their corresponding matrices.

Of course, as you well know, with matrix multiplication you have to be careful with the ordering. That is to say, in general, $AB \neq BA$. This can be put in more elegant terms by defining the **commutator**

$$[A, B] = AB - BA. \tag{1.21}$$

When $[A, B] \neq 0$ we then say the two operators do not commute. Commutators appear all the time. The commutation relations of a given set of operators is called the **algebra** of that set. And the algebra defines all properties of an operator. So in order to specify a physical theory, essentially all we need is the underlying algebra. We will see how that appears when we work out specific examples.

Commutators appear so often that it is useful to memorize the following formula:

$$[AB, C] = A[B, C] + [A, C]B \tag{1.22}$$

This formula is really easy to remember: first A goes out to the left then B goes out to the right. A similar formula holds for $[A, BC]$. Then B exists to the left and C exists to the right.

1.5 Eigenvalues and eigenvectors

When an operator acts on a vector, it produces another vector. But if you get lucky the operator may act on a vector and produce the same vector, up to a constant. When that happens, we say this vector is an eigenvector and the constant in front is the eigenvalue. In symbols,

$$A|\lambda = \lambda|\lambda\rangle. \tag{1.23}$$

The eigenvalues are the numbers λ and $|\lambda\rangle$ is the eigenvector associated with the eigenvalue λ.

Determining the structure of the eigenvalues and eigenvectors for an arbitrary operator may be a difficult task. One class of operators that is super well behaved are the **Hermitian operators**. Given an operator A, we define its adjoint as the operator $A^\dagger$ whose matrix elements are

$$(A^\dagger)_{i,j} = A^*_{j,i} \tag{1.24}$$

That is, we transpose and then take the complex conjugate. An operator is then said to be Hermitian when $A^\dagger = A$. Projection operators, for instance, are Hermitian.

The eigenvalues and eigenvectors of Hermitian operators are all well behaved and predictable:

1. Every Hermitian operator of dimension d always has d (not necessarily distinct) **real** eigenvalues.
2. The corresponding d eigenvectors can always be chosen to form an orthonormal basis.

An example of a Hermitian operator is the rank-1 projector $P_i = |i\rangle\langle i|$. It has one eigenvalue $\lambda = 1$ and all other eigenvalues zero. The eigenvector corresponding to $\lambda = 1$ is precisely $|i\rangle$ and the other eigenvectors are arbitrary combinations of the other basis vectors.

I will not prove the above properties, since they can be found in any linear algebra textbook or on Wikipedia. The proof that the eigenvalues are real, however, is cute and simple, so we can do it. Multiply Eq. (1.23) by $\langle\lambda|$, which gives

$$\langle\lambda|A|\lambda\rangle = \lambda. \tag{1.25}$$

Because of the relation (1.24), it now follows for any state that,

$$\langle\psi|A|\phi\rangle = \langle\phi|A^\dagger|\psi\rangle^*. \tag{1.26}$$

Taking the complex conjugate of Eq. (1.25) then gives

$$\langle\lambda|A^\dagger|\lambda\rangle = \lambda^*.$$

If $A^\dagger = A$ then we immediately see that $\lambda^* = \lambda$, so the eigenvalues are real.

Since the eigenvectors $|\lambda\rangle$ form a basis, we can decompose an operator A as in (1.19), but using the basis λ. We then get

$$A = \sum_\lambda \lambda|\lambda\rangle\langle\lambda|. \tag{1.27}$$

Thus, an operator A is diagonal when written in its own basis. That is why the procedure for finding eigenvalues and eigenvectors is called **diagonalization**.

1.6 Unitary matrices

A matrix U is called unitary when it satisfies:

$$UU^\dagger = U^\dagger U = 1, \tag{1.28}$$

Unitary matrices play a pivotal role in quantum mechanics. One of the main reasons for this is that they preserve the normalization of vectors. That is, if $|\psi'\rangle = U|\psi\rangle$ then $\langle\psi'|\psi'\rangle = \langle\psi|\psi\rangle$. Unitaries are the complex version of **rotation matrices**: when you rotate a vector, you don't change its magnitude, just the direction. The idea is exactly the same, except it is in $\mathbb{C}^d$ instead of $\mathbb{R}^3$.

Unitary matrices also appear naturally in the diagonalization of Hermitian operators that we just discussed [Eq. (1.23)]. Given the set of d eigenvectors $|\lambda_i\rangle$, define the matrix

$$U = \sum_i |\lambda_i\rangle\langle i|. \tag{1.29}$$

One can readily verify that since both $|i\rangle$ and $|\lambda_i\rangle$ form basis sets, this matrix will be unitary. The entries of U in the basis $|i\rangle$, $U_{ij} = \langle i|U|j\rangle$, are such that the **eigenvectors $|\lambda_i\rangle$ are arranged one in each column**:

$$U = \begin{pmatrix} \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ |\lambda_0\rangle & |\lambda_1\rangle & \dots & |\lambda_{d-1}\rangle \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \end{pmatrix} \tag{1.30}$$

This is one of those things that you simply need to stop and check for yourself. Please take a second to do that.

Next w apply this matrix U to the operator A:

$$U^\dagger A U = \sum_{i,j} |i\rangle\langle\lambda_i|A|\lambda_j\rangle\langle j| = \sum_i \lambda_i |i\rangle\langle i|.$$

Thus, we see that $U^\dagger A U$ produces a diagonal matrix with the eigenvalues λ_i. This is why finding the eigenstuff is called *diagonalization*. If we define

$$\Lambda = \sum_i \lambda_i |i\rangle\langle i| = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{d-1}) \tag{1.31}$$

Then, the *diagonal form* of A can be written as

$$\boxed{A = U\Lambda U^\dagger.} \tag{1.32}$$

Any Hermitian matrix may thus be diagonalized by a Unitary transformation.

1.7 Projective measurements and expectation values

As you know, in quantum mechanics measuring a system causes the wave-function to collapse. The simplest way of modeling this (which we will later generalize) is called a projective measurement. Let $|\psi\rangle$ be the state of the system at any given time. The postulate then goes as follows: If we measure in a certain basis $\{|i\rangle\}$, we will find the system in a given element $|i\rangle$ with probability

$$p_i = |\langle i|\psi\rangle|^2 \tag{1.33}$$

Moreover, if the system was found in state $|i\rangle$, then due to the action of the measurement the state after the measurement collapses to the state $|i\rangle$. That is, the measurement transforms the state as $|\psi\rangle \to |i\rangle$. I will not try to explain the physics behind this process right now. We will do that later one, in quite some detail. For now, just please accept that this crazy measurement thingy is actually possible.

The quantity $\langle i|\psi\rangle$ is the *probability amplitude* to find the system in $|i\rangle$. The modulus squared of the probability amplitude is the actual probability. The probabilities (1.33) are clearly non-negative. Moreover, they will sum to 1 when the state $|\psi\rangle$ is properly normalized:

$$\sum_i p_i = \sum_i \langle\psi|i\rangle\langle i|\psi\rangle = \langle\psi|\psi\rangle = 1.$$

This is why we introduced Eq. (1.6) back then.

Now let A be a Hermitian operator with eigenstuff $|\lambda_i\rangle$ and λ_i. If we measure in the basis $|\lambda_i\rangle$ then we can say that, with probability p_i the operator A was found in the eigenvalue λ_i. This is the idea of *measuring an observable*: we say an observable (Hermitian operator) can take on a set of values given by its eigenvalues λ_i, each occurring with probability $p_i = |\langle\lambda_i|\psi\rangle|^2$. Since any basis set $\{|i\rangle\}$ can always be associated with some observable, measuring in a basis or measuring an observable is actually the same thing.

Following this idea, we can also define the **expectation value** of the operator A. But to do that, we must define it as an *ensemble average*. That is, we prepare many identical copies of our system and then measure each copy, discarding it afterwards. If we measure the same system sequentially, we will just obtain the same result over and over again, since in a measurement we collapsed the state.[1] From the data we collect, we construct the probabilities p_i. The expectation value of A will then be

$$\langle A\rangle := \sum_i \lambda_i p_i \tag{1.34}$$

I will leave for you to show that using Eq. (1.33) we may also write this as

$$\langle A\rangle := \langle\psi|A|\psi\rangle \tag{1.35}$$

The expectation value of the operator is therefore the **sandwich** (yummmm) of A on $|\psi\rangle$.

The word "projective" in projective measurement also becomes clearer if we define the projection operators $P_i = |i\rangle\langle i|$. Then the probabilities (1.33) become

$$p_i = \langle\psi|P_i|\psi\rangle. \tag{1.36}$$

The probabilities are therefore nothing but the expectation value of the projection operators on the state $|\psi\rangle$.

---

[1] To be more precise, after we collapse, the state will start to evolve in time. If the second measurement occurs right after the first, nothing will happen. But if it takes some time, we may get something non-trivial. We can also keep on measuring a system on purpose, to always push it to a given a state. That is called the Zeno effect.

1.8 Pauli matrices

As far as qubits are concerned, the most important matrices are the Pauli matrices. They are defined as

$$\boxed{\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \tag{1.37}$$

The Pauli matrices are both Hermitian, $\sigma_i^\dagger = \sigma_i$ and unitary, $\sigma_i^2 = 1$. The operator σ_z is diagonal in the $|0\rangle, |1\rangle$ basis:

$$\sigma_z|0\rangle = |0\rangle, \qquad \sigma_z|1\rangle = -|1\rangle. \tag{1.38}$$

The operators σ_x and σ_y, on the other hand, flip the qubit. For instance,

$$\sigma_x|0\rangle = |1\rangle, \qquad \sigma_x|1\rangle = |0\rangle. \tag{1.39}$$

The action of σ_y is similar, but gives a factor of $\pm i$ depending on the flip.

Another set of operators that are commonly used are the **lowering and raising operators**:

$$\boxed{\sigma_+ = |0\rangle\langle 1| = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \sigma_- = |1\rangle\langle 0| = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}} \tag{1.40}$$

They are related to $\sigma_{x,y}$ according to

$$\sigma_x = \sigma_+ + \sigma_- \qquad \text{and} \qquad \sigma_y = -i(\sigma_+ - \sigma_-) \tag{1.41}$$

or

$$\sigma_\pm = \frac{\sigma_x \pm i\sigma_y}{2} \tag{1.42}$$

The action of these operators on the states $|0\rangle$ and $|1\rangle$ can be a bit counter-intuitive:

$$\sigma_+|1\rangle = |0\rangle, \qquad \text{and} \qquad \sigma_-|0\rangle = |1\rangle \tag{1.43}$$

In the way we defined the Pauli matrices, the indices x, y and z may seem rather arbitrary. They acquire a stronger physical meaning in terms of Bloch's sphere. The states $|0\rangle$ and $|1\rangle$ are eigenstates of σ_z and they lie along the z axis of the Bloch sphere. Similarly, the states $|\pm\rangle$ in Eq. (1.11) can be verified to be eigenstates of σ_x and they lie on the x axis. One can generalize this and define the Pauli matrix in an arbitrary direction of Bloch's sphere by first defining a unit vector

$$\boldsymbol{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{1.44}$$

where $\theta \in [0, \pi)$ and $\phi \in [0, 2\pi]$. The spin operator at an arbitrary direction $\boldsymbol{n}$ is then defined as

$$\sigma_{\boldsymbol{n}} = \boldsymbol{\sigma} \cdot \boldsymbol{n} = \sigma_x n_x + \sigma_y n_y + \sigma_z n_z \tag{1.45}$$

Please take a second to check that we can recover $\sigma_{x,y,z}$ just by taking appropriate choices of θ and ϕ. In terms of the parametrization (1.44) this spin operator becomes

$$\sigma_{\boldsymbol{n}} = \begin{pmatrix} n_z & n_x - i n_y \\ n_x + i n_y & -n_z \end{pmatrix} = \begin{pmatrix} \cos\theta & e^{-i\phi}\sin\theta \\ e^{i\phi}\sin\theta & -\cos\theta \end{pmatrix} \tag{1.46}$$

I will leave for you the exercise of computing the eigenvalues and eigenvectors of this operator. The eigenvalues are ± 1, which is quite reasonable from a physical perspective since the eigenvalues are a property of the operator and thus should not depend on our choice of orientation in space. In other words, the spin components in any direction in space are always ± 1. As for the eigenvectors, they are

$$|\boldsymbol{n}_+\rangle = \begin{pmatrix} e^{-i\phi/2}\cos\frac{\theta}{2} \\ e^{i\phi/2}\sin\frac{\theta}{2} \end{pmatrix}, \qquad |\boldsymbol{n}_-\rangle = \begin{pmatrix} -e^{-i\phi/2}\sin\frac{\theta}{2} \\ e^{i\phi/2}\cos\frac{\theta}{2} \end{pmatrix} \tag{1.47}$$

If we stare at this for a second, then the connection with Bloch's sphere in Fig. 1.1 starts to appear: the state $|\boldsymbol{n}_+\rangle$ is exactly the same as the Bloch sphere parametrization (1.10), except for a global phase $e^{-i\phi/2}$. Moreover, the state $|\boldsymbol{n}_-\rangle$ is simply the state opposite to $|\boldsymbol{n}_+\rangle$.

Another connection to Bloch's sphere is obtained by computing the expectation values of the spin operators in the state $|\boldsymbol{n}_+\rangle$. They read

$$\langle\sigma_x\rangle = \sin\theta\cos\phi, \qquad \langle\sigma_y\rangle = \sin\theta\sin\phi, \qquad \langle\sigma_z\rangle = \cos\theta \tag{1.48}$$

Thus, the average of σ_i is simply the i-th component of $\boldsymbol{n}$: it makes sense! We have now gone full circle: we started with $\mathbb{C}^2$ and made a parametrization in terms of a unit sphere in $\mathbb{R}^3$. Now we defined a point $\boldsymbol{n}$ in $\mathbb{R}^3$, as in Eq. (1.44), and showed how to write the corresponding state in $\mathbb{C}^2$, Eq. (1.47).

To finish, let us also write the diagonalization of $\sigma_{\boldsymbol{n}}$ in the form of Eq. (1.32). To do that, we construct a matrix whose columns are the eigenvectors $|\boldsymbol{n}_+\rangle$ and $|\boldsymbol{n}_-\rangle$. This matrix is then

$$G = \begin{pmatrix} e^{-i\phi/2}\cos\frac{\theta}{2} & -e^{-i\phi/2}\sin\frac{\theta}{2} \\ e^{i\phi/2}\sin\frac{\theta}{2} & e^{i\phi/2}\cos\frac{\theta}{2} \end{pmatrix} \tag{1.49}$$

The diagonal matrix Λ in Eq. (1.32) is the matrix containing the eigenvalues ± 1. Hence it is precisely σ_z. Thus, we conclude that

$$\sigma_{\boldsymbol{n}} = G\sigma_z G^\dagger \tag{1.50}$$

We therefore see that G is the unitary matrix that "rotates" a spin operator from an arbitrary direction towards the z direction.

The Pauli matrices can also be used as a mathematical trick to simplify some calculations with general 2×2 matrices. Finding eigenvalues is easy. But surprisingly, the eigenvectors can become quite ugly. I way to circumvent this is to express them in terms of Pauli matrices σ_x, σ_y, σ_z and $\sigma_0 = 1$ (the identity matrix). We can write this in an organized way as

$$A = a_0 + \boldsymbol{a} \cdot \boldsymbol{\sigma}, \tag{1.51}$$

for a certain set of four numbers a_0, a_x, a_y and a_z. Next define $a = |\boldsymbol{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$ and $\boldsymbol{n} = \boldsymbol{a}/a$. Then A can be written as

$$A = a_0 + a(\boldsymbol{n} \cdot \boldsymbol{\sigma}) \tag{1.52}$$

The following silly properties of eigenthings are now worth remembering:

1. If $A|\lambda\rangle = \lambda|\lambda\rangle$ and $B = \alpha A$ then the eigenvalues of B will be $\lambda_B = \alpha\lambda$.
2. If $A|\lambda\rangle = \lambda|\lambda\rangle$ and $B = A + c$ then the eigenvalues of B will be $\lambda_B = \lambda + c$.

Moreover, in both cases, the eigenvectors of B are the same as those of A. Looking at Eq. (1.52), we then see that

$$\text{eigs}(A) = a_0 \pm a \tag{1.53}$$

As for the eigenvectors, they will be given precisely by Eq. (1.47), where the angles θ and ϕ are defined in terms of the unit vector $\boldsymbol{n} = \boldsymbol{a}/a$. Thus, we finally conclude that any 2 × 2 matrix may be diagonalized as

$$A = G(a_0 + a\sigma_z)G^\dagger \tag{1.54}$$

This gives an elegant way of writing the eigenvectors of 2 × 2 matrices.

1.9 Functions of operators

Let $f(x)$ be an arbitrary function, such as e^x or $1/(1-x)$, etc. We consider here functions that have a Taylor series expansion $f(x) = \sum_n c_n x^n$, for some coefficients c_n. We define the application of this function to an operator, $f(A)$, *by means of the Taylor series*. That is, we define

$$f(A) := \sum_{n=0}^{\infty} c_n A^n \tag{1.55}$$

Functions of operators are super easy to compute when an operator is diagonal. Consider, for instance, a Hermitian operator A decomposed as $A = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$. Since $\langle\lambda_i|\lambda_j\rangle = \delta_{i,j}$ it follows that

$$A^2 = \sum_i \lambda_i^2 |\lambda_i\rangle\langle\lambda_i|.$$

Thus, we see that the eigenvalues of A^2 are λ_i^2, whereas the eigenvectors are the same as those of A. Of course, this is also true for A^3 or any other power. Inserting this in Eq. (1.55) we then get

$$f(A) = \sum_{n=0}^{\infty} c_n \sum_i \lambda_i^n |\lambda_i\rangle\langle\lambda_i| = \sum_i \bigg[\sum_n c_n \lambda_i^n\bigg] |\lambda_i\rangle\langle\lambda_i|.$$

The quantity inside the square brackets is nothing but $f(\lambda_i)$. Thus we conclude that

$$\boxed{f(A) = \sum_i f(\lambda_i)|\lambda_i\rangle\langle\lambda_i|} \tag{1.56}$$

This result is super useful: if an operator is diagonal, simply apply the function to the diagonal elements, like one would do for numbers. For example, the exponential of the Pauli matrix σ_z in Eq. (1.37) simply reads

$$e^{-i\phi\sigma_z/2} = \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix}. \tag{1.57}$$

When the matrix is not diagonal, a bit more effort is required, as we now discuss.

The infiltration property of unitaries

Next suppose we have two operators, A and B, which are connected by means of an arbitrary unitary transformation, $A = UBU^\dagger$. It then follows that

$$A^2 = (UBU^\dagger)(UBU^\dagger) = UB(U^\dagger U)BU^\dagger = UB^2U^\dagger.$$

The unitaries in the middle cancel and we get the same structure as $A = UBU^\dagger$, but for A^2 and B^2. Similarly, if we continue like this we will get $A^3 = UB^3U^\dagger$ or, more generally, $A^n = UB^nU^\dagger$. Plugging this in Eq. (1.55) then yields

$$f(A) = \sum_{n=0}^\infty c_n UB^nU^\dagger = U\bigg[\sum_{n=0}^\infty c_n B^n\bigg]U^\dagger.$$

Thus, we reach the remarkable conclusion that

$$\boxed{f(A) = f(UBU^\dagger) = Uf(B)U^\dagger.} \tag{1.58}$$

I call this the **infiltration property** of the unitary. Unitaries are sneaky! They can enter inside functions, as long as you have U on one side and $U^\dagger$ in the other. This is the magic of unitaries. Right here, in front of your eyes. This is why unitaries are so remarkable.

From this result we also get a simple recipe on how to compute the matrix exponentials of general, non-diagonal matrices. Simply write A in diagonal form as $A = U\Lambda U^\dagger$, where Λ is the diagonal matrix. Then, computing $f(\Lambda)$ is easy because all we need is to apply the function to the diagonal entries. After we do that, we simply multiple by $U(\dot{})U^\dagger$ to get back the full matrix. As an example, consider a general 2×2 matrix with the eigenstructure given by Eq. (1.54). One then finds that

$$e^{i\theta A} = G\begin{pmatrix} e^{i\theta(a_0+a)} & 0 \\ 0 & e^{i\theta(a_0-a)} \end{pmatrix}G^\dagger. \tag{1.59}$$

Of course, if we carry out the full multiplications, this will turn out to be a bit ugly, but still, it is a closed formula for the exponential of an arbitrary 2×2 matrix. We can also write down the same formula for the inverse:

$$A^{-1} = G\begin{pmatrix} \frac{1}{a_0+a} & 0 \\ 0 & \frac{1}{a_0-a} \end{pmatrix} G^\dagger. \tag{1.60}$$

After all, inverse is also a function of a matrix.

More about exponentials

The most important function by far the exponential of an operator, defined by the Taylor series

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots, \tag{1.61}$$

Using our two basic formulas (1.56) and (**??**) we then get

$$e^A = \sum_i e^{\lambda_i} |\lambda_i\rangle\langle\lambda_i| = U e^\Lambda U^\dagger \tag{1.62}$$

To practice, let us compute the exponential of some other Pauli operators. The eigenvectors of σ_x, for instance, are the $|\pm\rangle$ states in Eq. (1.11). Thus

$$e^{i\alpha\sigma_x} = e^{i\alpha}|+\rangle\langle+| + e^{-i\alpha}|-\rangle\langle-| = \begin{pmatrix} \cos\alpha & i\sin\alpha \\ i\sin\alpha & \cos\alpha \end{pmatrix} = \cos\alpha + i\sigma_x \sin\alpha \tag{1.63}$$

It is also interesting to compute this in another way. Recall that $\sigma_x^2 = 1$. In fact, this is true for any Pauli matrix $\sigma_{\boldsymbol{n}}$. We can use this to compute $e^{i\alpha\sigma_{\boldsymbol{n}}}$ via the definition of the exponential in Eq. (1.61). Collecting the terms proportional to $\sigma_{\boldsymbol{n}}$ and $\sigma_{\boldsymbol{n}}^2 = 1$ we get:

$$e^{i\alpha\sigma_{\boldsymbol{n}}} = \left[1 - \frac{\alpha^2}{2} + \frac{\alpha^4}{4!} + \ldots\right] + \sigma_{\boldsymbol{n}}\left[i\alpha - i\frac{\alpha^3}{3!} + \ldots\right].$$

Thus, we readily see that

$$e^{i\alpha\sigma_{\boldsymbol{n}}} = \cos\alpha + i\sigma_{\boldsymbol{n}} \sin\alpha, \tag{1.64}$$

where I remind you that the first term in Eq. (1.64) is actually $\cos\alpha$ multiplying the identity matrix. If we now replace $\sigma_{\boldsymbol{n}}$ by σ_x, we recover Eq. (1.63). It is interesting to point out that nowhere did we use the fact that the matrix was 2×2. If you are ever given a matrix, of arbitrary dimension, but such that $A^2 = 1$, then the same result will also apply.

In the theory of angular momentum, we learn that the operator which affects a rotation around a given axis, defined by a vector $\boldsymbol{n}$, is given by $e^{-i\alpha\sigma_{\boldsymbol{n}}/2}$. We can use this to construct the state $|\boldsymbol{n}_+\rangle$ in Eq. (1.47). If we start in the north pole, we can get to a general point in the $\mathbb{R}^3$ unit sphere by two rotations. First you rotate around the y axis by an angle θ and then around the z axis by an angle ϕ (take a second to imagine how this works in your head). Thus, one would expect that

$$|\boldsymbol{n}_+\rangle = e^{-i\phi\sigma_z/2} e^{-i\theta\sigma_y/2} |0\rangle. \tag{1.65}$$

I will leave for you to check that this is indeed Eq. (1.47). Specially in the context of more general spin operators, these states are also called **spin coherent states**, since they are the closest analog to a point in the sphere. The matrix G in Eq. (1.49) can also be shown to be

$$G = e^{-i\phi\sigma_z/2} e^{-i\theta\sigma_y/2} \tag{1.66}$$

The exponential of an operator is defined by means of the Taylor series (1.61). However, that does not mean that it behaves just like the exponential of numbers. In fact, the exponential of an operator *does not* satisfy the exponential property:

$$e^{A+B} \neq e^A e^B. \tag{1.67}$$

In a sense this is obvious: the left-hand side is symmetric with respect to exchanging A and B, whereas the right-hand side is not since e^A does not necessarily commute with e^B. Another way to see this is by means of the interpretation of $e^{i\alpha\sigma_n}$ as a rotation: rotations between different axes do not in general commute.

Exponentials of operators is a serious business. There is a vast mathematical literature on dealing with them. In particular, there are a series of popular formulas which go by the generic name of Baker-Campbell-Hausdorff (BCH) formulas. For instance, there is a BCH formula for dealing with e^{A+B}, which in Wikipedia is also called Zassenhaus formula. It reads

$$e^{t(A+B)} = e^{tA} e^{tB} e^{-\frac{t^2}{2}[A,B]} e^{\frac{t^3}{3!}(2[B,[A,B]]+[A,[A,B]])} \ldots, \tag{1.68}$$

where t is just a parameter to help keep track of the order of the terms. From the fourth order onwards, things just become mayhem. There is really no mystery behind this formula: it simply summarizes the *ordering* of non-commuting objects. You can derive it by expanding both sides in a Taylor series and grouping terms of the same order in t. It is a really annoying job, so everyone just trusts Zassenhaus. Notwithstanding, we can extract some physics out of this. In particular, suppose t is a tiny parameter. Then Eq. (1.68) can be seen as a series expansion in t: the error you make in writing $e^{t(A+B)}$ as $e^{tA}e^{tB}$ will be a term proportional to t^2. A particularly important case of Eq. (1.68) is when $[A, B]$ commutes with both A and B. That generally means $[A, B] = c$, a number. But it can also be that $[A, B]$ is just some fancy matrix which happens to commute with both A and B. We see in Eq. (1.68) that in this case all higher order terms commute and the series truncates. That is

$$e^{t(A+B)} = e^{tA} e^{tB} e^{-\frac{t^2}{2}[A,B]}, \qquad \text{when } [A,[A,B]] = 0 \text{ and } [B,[A,B]] = 0 \tag{1.69}$$

There is also another BCH formula that is very useful. It deals with the sandwich of an operator between two exponentials, and reads

$$e^{tA} B e^{-tA} = B + t[A,B] + \frac{t^2}{2!}[A,[A,B]] + \frac{t^3}{3!}[A,[A,[A,B]]] + \ldots \tag{1.70}$$

Again, you can derive this formula by simply expanding the left-hand side and collecting terms of the same order in t. I suggest you give it a try in this case, at least up to order t^2. That will help give you a feeling of how messy things can get when dealing with non-commuting objects.

1.10 The Trace

The trace of an operator is defined as the sum of its diagonal entries:

$$\operatorname{tr}(A) = \sum_i \langle i|A|i\rangle. \tag{1.71}$$

It turns out that the trace is the same no matter which basis you use. You can see that using completeness: for instance, if $|a\rangle$ is some other basis then

$$\sum_i \langle i|A|i\rangle = \sum_i \sum_a \langle i|a\rangle\langle a|A|i\rangle = \sum_i \sum_a \langle a|A|i\rangle\langle i|a\rangle = \sum_a \langle a|A|a\rangle.$$

Thus, we conclude that

$$\operatorname{tr}(A) = \sum_i \langle i|A|i\rangle = \sum_a \langle a|A|a\rangle. \tag{1.72}$$

The trace is a property of the operator, not of the basis you choose. Since it does not matter which basis you use, let us choose the basis $|\lambda_i\rangle$ which diagonalizes the operator A. Then $\langle \lambda_i|A|\lambda_i\rangle = \lambda_i$ will be an eigenvalue of A. Thus, we also see that

$$\boxed{\operatorname{tr}(A) = \sum_i \lambda_i = \text{sum of all eigenvalues of } A}. \tag{1.73}$$

Perhaps the most useful property of the trace is that it is *cyclic*:

$$\operatorname{tr}(AB) = \operatorname{tr}(BA). \tag{1.74}$$

I will leave it for you to demonstrate this. Simply insert a convenient completeness relation in the middle of AB. Using the cyclic property (1.74) you can also move around an arbitrary number of operators, but only in cyclic permutations. For instance:

$$\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA). \tag{1.75}$$

Note how I am moving them around in a specific order: $\operatorname{tr}(ABC) \neq \operatorname{tr}(BAC)$. An example that appears often is a trace of the form $\operatorname{tr}(UAU^\dagger)$, where U is unitary operator. In this case, it follows from the cyclic property that

$$\operatorname{tr}(UAU^\dagger) = \operatorname{tr}(AU^\dagger U) = \operatorname{tr}(A)$$

Thus, the trace of an operator is invariant by unitary transformations. This is also in line with the fact that the trace is the sum of the eigenvalues and unitaries preserve eigenvalues.

Finally, let $|\psi\rangle$ and $|\phi\rangle$ be arbitrary kets and let us compute the trace of the outer product $|\psi\rangle\langle\phi|$:

$$\operatorname{tr}(|\psi\rangle\langle\phi|) = \sum_i \langle i|\psi\rangle\langle\phi|i\rangle = \sum_i \langle\phi|i\rangle\langle i|\psi\rangle$$

The sum over $|i\rangle$ becomes a 1 due to completeness and we conclude that

$$\boxed{\operatorname{tr}(|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle.} \tag{1.76}$$

Notice how this follows the same logic as Eq. (1.74), so you can pretend you just used the cyclic property. This formula turns out to be extremely useful, so it is definitely worth remembering.

Chapter 2

Density matrix theory

2.1 The density matrix

A ket $|\psi\rangle$ is actually not the most general way of defining a quantum state. To motivate this, consider the state $|\boldsymbol{n}_+\rangle$ in Eq. (1.47) and the corresponding expectation values computed in Eq. (1.48). This state is always poiting *somewhere*: it points at the direction $\boldsymbol{n}$ of the Bloch sphere. It is impossible, for instance, to find a quantum ket which is isotropic. That is, where $\langle\sigma_x\rangle = \langle\sigma_y\rangle = \langle\sigma_z\rangle = 0$. That sounds strange. The solution to this conundrum lies in the fact that we need to also introduce some *classical uncertainty* to the problem. Kets are only able to encompass quantum uncertainty.

The most general representation of a quantum system is written in terms of an operator ρ called the *density operator*, or *density matrix*. It is built in such a way that it naturally encompasses both quantum and classical probabilities. But that is not all. We will also learn next chapter that density matrices are intimately related to **entanglement**. So even if we have *no* classical uncertainties, we will also eventually find the need for dealing with density matrices. For this reason, the density matrix is *the most important concept in quantum theory*. I am not exaggerating. You started this chapter as a kid. You will finish it as an adult. :)

To motivate the idea, imagine we have a machine which prepares quantum systems in certain states. For instance, this could be an oven producing spin 1/2 particles, or a quantum optics setup producing photons. But suppose that this apparatus is imperfect, so it does not always produces the same state. That is, suppose that it produces a state $|\psi_1\rangle$ with a certian probability p_1 or a state $|\psi_2\rangle$ with a certain probability p_2 and so on. Notice how we are introducing here a *classical uncertainty*. The $|\psi_i\rangle$ are quantum states, but we simply *don't know* which states we will get out of the machine. We can have as many p's as we want. All we need to assume is that satisfy the properties expected from a probability:

$$p_i \in [0,1], \qquad \text{and} \qquad \sum_i p_i = 1 \tag{2.1}$$

Now let A be an observable. If the state is $|\psi_1\rangle$, then the expectation value of A will be $\langle\psi_1|A|\psi_1\rangle$. But if it is $|\psi_2\rangle$ then it will be $\langle\psi_2|A|\psi_2\rangle$. To compute the actual expectation value of A we must therefore perform an **average of quantum averages**:

$$\langle A\rangle = \sum_i p_i \langle\psi_i|A|\psi_i\rangle \tag{2.2}$$

We simply weight the possible expectation values $\langle\psi_i|A|\psi_i\rangle$ by the relative probabilities p_i that each one occurs.

What is important to realize is that this type of average *cannot* be writen as $\langle\phi|A|\phi\rangle$ for some ket $|\phi\rangle$. If we want to attribute a "state" to our system, then we must generalize the idea of ket. To do that, we use Eq. (1.76) to write

$$\langle\psi_i|A|\psi_i\rangle = \mathrm{tr}\Big[A|\psi_i\rangle\langle\psi_i|\Big]$$

Then Eq. (2.2) may be written as

$$\langle A\rangle = \sum_i p_i\, \mathrm{tr}\Big[A|\psi_i\rangle\langle\psi_i|\Big] = \mathrm{tr}\bigg\{A\sum_i p_i|\psi_i\rangle\langle\psi_i|\bigg\}$$

This motivates us to define the **density matrix** as

$$\boxed{\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|} \tag{2.3}$$

Then we may finally write Eq. (2.2) as

$$\boxed{\langle A\rangle = \mathrm{tr}(A\rho)} \tag{2.4}$$

which, by the way, is the same as $\mathrm{tr}(\rho A)$ since the trace is cyclic [Eq. (1.74)].

With this idea, we may now recast *all* of quantum mechanics in terms of density matrices, instead of kets. If it happens that a density matrix can be written as $\rho = |\psi\rangle\langle\psi|$, we say we have a **pure state**. And in this case it is not necessary to use ρ at all. One may simply continue to use $|\psi\rangle$. For instance, Eq. (2.4) reduces to the usual result: $\mathrm{tr}(A\rho) = \langle\psi|A|\psi\rangle$. A state which is not pure is usually called a **mixed state**. In this case kets won't do us no good and we *must* use ρ.

Examples

Let' s play with some examples. To start, suppose a machine tries to produce qubits in the state $|0\rangle$. But it is not very good so it only produces $|0\rangle$ with probability p. And, with probability $1-p$ it produces the state $|1\rangle$. The density matrix would then be.

$$\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = \begin{pmatrix} p & 0 \\ 0 & 1-p\end{pmatrix}.$$

Or it could be such that it produces either $|0\rangle$ or $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Then,

$$\rho = p|0\rangle\langle 0| + (1-p)|+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1+p & 1-p \\ 1-p & 1-p \end{pmatrix}.$$

Maybe if our device is not completely terrible, it will produce most of the time $|0\rangle$ and every once in a while, a state $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle$, where θ is some small angle. The density matrix for this system will then be

$$\rho = p|0\rangle\langle 0| + (1-p)|\psi\rangle\langle\psi| = \begin{pmatrix} p + (1-p)\cos^2\frac{\theta}{2} & (1-p)\sin\frac{\theta}{2}\cos\frac{\theta}{2} \\ (1-p)\sin\frac{\theta}{2}\cos\frac{\theta}{2} & (1-p)\sin^2\frac{\theta}{2} \end{pmatrix}$$

Of course, the machine can very well produce more than 2 states. But you get the idea.

Next let's talk about something really cool (and actually quite deep), called the **ambiguity of mixtures**. The idea is quite simple: if you mix stuff, you generally loose information, so you don't always know where you started at. To see what I mean, consider a state which is a 50-50 mixture of $|0\rangle$ and $|1\rangle$. The corresponding density matrix will then be

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Alternatively, consider a 50-50 mixture of the states $|\pm\rangle$ in Eq. (1.11). In this case we get

$$\rho = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that both are identical. Hence, we have *no way to tell* if we began with a 50-50 mixture of $|0\rangle$ and $|1\rangle$ or of $|+\rangle$ and $|-\rangle$. By mixing stuff, we have lost information.

2.2 Properties of the density matrix

The density matrix satisfies a bunch of very special properties. We can figure them out using only the definition (2.3) and recalling that $p_i \in [0,1]$ and $\sum_i p_i = 1$ [Eq. (2.1)]. First, the density matrix is a Hermitian operator:

$$\rho^\dagger = \rho. \tag{2.5}$$

Second,

$$\text{tr}(\rho) = \sum_i p_i \,\text{tr}(|\psi_i\rangle\langle\psi_i|) = \sum_i p_i \langle\psi_i|\psi_i\rangle = \sum_i p_i = 1. \tag{2.6}$$

This is the *normalization* condition of the density matrix. Another way to see this is from Eq. (2.4) by choosing $A = 1$. Then, since $\langle 1\rangle = 1$ we again get $\text{tr}(\rho) = 1$.

We also see from Eq. (2.8) that $\langle\phi|\rho|\phi\rangle$ is a sum of quantum probabilities $|\langle\phi|\psi_i\rangle|^2$ averaged by classical probabilities p_i. This entails the following interpretation: for an arbitrary state $|\phi\rangle$,

$$\boxed{\langle\phi|\rho|\phi\rangle = \text{Prob. of finding the system at state } |\phi\rangle \text{ given that it's state is } \rho} \tag{2.7}$$

Besides normalization, the other big property of a density matrix is that it is **positive semi-definite**, which we write symbolically as $\rho \geq 0$. What this means is that *its sandwich in any quantum state is always non-negative*. In symbols, if $|\phi\rangle$ is an arbitrary quantum state then

$$\langle\phi|\rho|\phi\rangle = \sum_i p_i |\langle\phi|\psi_i\rangle|^2 \geq 0. \tag{2.8}$$

Of course, this makes sense in view of the probabilistic interpretation of Eq. (2.7). Please note that this does *not* mean that all entries of ρ are non-negative. Some of them may be negative. It does mean, however, that the diagonal entries are always non-negative, no matter which basis you use.

Another equivalent definition of a positive semi-definite operator is one whose **eigenvalues are always non-negative**. In Eq. (2.3) it already looks as if ρ is in diagonal form. However, we need to be a bit careful because the $|\psi_i\rangle$ are arbitrary states and do not necessarily form a basis (which can be seen explicitly in the examples given above). Thus, in general, the diagonal structure of ρ will be different. Notwithstanding, ρ is Hermitian and may therefore be diagonalized by some orthonormal basis $|\lambda_k\rangle$ as

$$\rho = \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|, \tag{2.9}$$

for certain eigenvalues λ_k. Since Eq. (2.8) must be true for any state $|\phi\rangle$ we may choose, in particular, $|\phi\rangle = |\lambda_k\rangle$, which gives

$$\lambda_k = \langle\lambda_k|\rho|k\rangle \geq 0.$$

Thus, we see that the statement of positive semi-definiteness is equivalent to saying that the eigenvalues are non-negative. In addition to this, we also have that $\text{tr}(\rho) = 1$, which implies that $\sum_k \lambda_k = 1$. Thus we conclude that the eigenvalues of ρ behave like probabilities:

$$\lambda_k \in [0,1], \qquad \sum_k \lambda_k = 1. \tag{2.10}$$

But they are not the same probabilities p_i. They just behave like a set of probabilities, that is all.

For future reference, let me summarize what we learned in a big box: the basic properties of a density matrix are

$$\boxed{\text{Defining properties of a density matrix:} \qquad \text{tr}(\rho) = 1 \quad \text{and} \quad \rho \geq 0.} \tag{2.11}$$

Any normalized positive semi-definite matrix is a valid candidate for a density matrix.

I emphasize again that the notation $\rho \geq 0$ in Eq. (2.11) means the matrix is positive semi-definite, not that the entries are positive. For future reference, let me list here some properties of positive semi-definite matrices:

- $\langle\phi|\rho|\phi\rangle \geq 0$ for any state $|\phi\rangle$;
- The eigenvalues of ρ are always non-negative.
- The diagonal entries are always non-negative.
- The off-diagonal entries in any basis satisfy $|\rho_{ij}| \leq \sqrt{\rho_{ii}\rho_{jj}}$.

2.3 Purity

Next let us look at ρ^2. The eigenvalues of this matrix are λ_k^2 so

$$\mathrm{tr}(\rho^2) = \sum_k \lambda_k^2 \leq 1 \tag{2.12}$$

The only case when $\mathrm{tr}(\rho^2) = 1$ is when ρ is a pure state. In that case it can be written as $\rho = |\psi\rangle\langle\psi|$ so it will have one eigenvalue $p_1 = 1$ and all other eigenvalues equal to zero. Hence, the quantity $\mathrm{tr}(\rho^2)$ represents the **purity** of the quantum state. When it is 1 the state is pure. Otherwise, it will be smaller than 1:

$$\boxed{\text{Purity} = \mathcal{P} := \mathrm{tr}(\rho^2) \leq 1} \tag{2.13}$$

As a side note, when the dimension of the Hilbert space d is finite, it also follows that $\mathrm{tr}(\rho^2)$ will have a lower bound:

$$\frac{1}{d} \leq \mathrm{tr}(\rho^2) \leq 1 \tag{2.14}$$

This lower bound occurs when ρ is the **maximally disordered state**

$$\rho = \frac{\mathbb{I}_d}{d} \tag{2.15}$$

where $\mathbb{I}_d$ is the identity matrix of dimension d.

2.4 Bloch's sphere and coherence

The density matrix for a qubit will be 2×2 and may therefore be parametrized as

$$\rho = \begin{pmatrix} p & q \\ q^* & 1-p \end{pmatrix}, \tag{2.16}$$

where $p \in [0, 1]$ and I used $1 - p$ in the last entry due to the normalization $\mathrm{tr}(\rho^2) = 1$. If the state is pure then it can be written as $|\psi\rangle = a|0\rangle + b|1\rangle$, in which case the density matrix becomes

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{pmatrix}. \tag{2.17}$$

This is the density matrix of a system which is in a *superposition* of $|0\rangle$ and $|1\rangle$. Conversely, we could construct a state which can be in $|0\rangle$ **or** $|1\rangle$ with different probabilities. According to the very definition of the density matrix in Eq. (2.3), this state would be

$$\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}. \tag{2.18}$$

This is a classical state, obtained from classical probability theory. The examples in Eqs. (2.17) and (2.18) reflect well the difference between quantum superpositions and classical probability distributions.

Another convenient way to write the state (2.16) is as

$$\boxed{\rho = \frac{1}{2}(1 + \boldsymbol{s} \cdot \boldsymbol{\sigma}) = \frac{1}{2}\begin{pmatrix} 1 + s_z & s_x - i s_y \\ s_x + i s_y & 1 - s_z \end{pmatrix}.} \tag{2.19}$$

where $\boldsymbol{s} = (s_x, s_y, s_z)$ is a vector. The physical interpretation of $\boldsymbol{s}$ becomes evident from the following relation:

$$s_i = \langle \sigma_i \rangle = \mathrm{tr}(\sigma_i \rho). \tag{2.20}$$

The relation between these parameters and the parametrization in Eq. (2.16) is

$$\begin{aligned} \langle \sigma_x \rangle &= q + q^*, \\ \langle \sigma_y \rangle &= i(q - q^*), \\ \langle \sigma_z \rangle &= 2p - 1. \end{aligned}$$

Next we look at the purity of a qubit density matrix. From Eq. (2.19) one readily finds that

$$\mathrm{tr}(\rho^2) = \frac{1}{2}(1 + \boldsymbol{s}^2). \tag{2.21}$$

Thus, due to Eq. (2.12), it also follows that

$$\boldsymbol{s}^2 = s_x^2 + s_y^2 + s_z^2 \leq 1. \tag{2.22}$$

When $\boldsymbol{s}^2 = 1$ we are in a pure state. In this case the vector $\boldsymbol{s}$ lies on the surface of the Bloch sphere. For mixed states $\boldsymbol{s}^2 < 1$ and the vector is inside the Bloch sphere. Thus, we see that the purity can be directly associated with the radius in the sphere. This is pretty cool! The smaller the radius, the more mixed is the state. In particular, the maximally disordered state occurs when $\boldsymbol{s} = 0$ and reads

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.23}$$

In this case the state lies in the center of the sphere. A graphical representation of pure and mixed states in the Bloch sphere is shown in Fig. 2.1.

2.5 Schrödinger and von Neumann

We will now talk about how states evolve in time. Kets evolve according to Schrödinger's equation. When Schrödinger's equation is written for density matrices, it then goes by the name of **von Neumann's equation**. However, as we will learn, von Neumann's equation is not the most general kind of quantum evolution, which is what we will call a Quantum Channel or **Quantum Operation**. The theory of quantum operations is *awesome*. Here I just want to give you a quick look at it, but we will get back to this many times again.

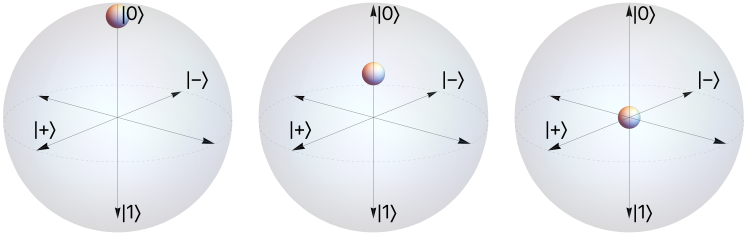

Figure 2.1: Examples of pure and mixed states in the z axis. Left: a pure state. Center: an arbitrary mixed state. Right: the maximally mixed state (2.23).

Schrödinger's equation

We start with Schödinger's equation. Interestingly, the structure of Schödinger's equation can be obtained by only postulating that the transformation caused by the time evolution should be a *linear* operation, in the sense that it corresponds to the action of a linear operator on the original state. That is, we can write the time evolution from time t_0 to time t as

$$|\psi(t)\rangle = U(t,t_0)|\psi(t_0)\rangle, \tag{2.24}$$

where $U(t,t_0)$ is the operator which affects the transformation between states. This assumption of linearity is one of the most fundamental properties of quantum mechanics and, in the end, is really based on experimental observations.

In addition to linearity, we continue to assume the probabilistic interpretation of kets, which mean they must remain normalized. That is, they must satisfy $\langle\psi(t)|\psi(t)\rangle = 1$ at all times. Looking at Eq. (2.24), we see that this will only be true when the matrix $U(t,t_0)$ is unitary. Hence, we conclude that *the time evolution must be described by a unitary matrix*. This is very important and, as already mentioned, is ultimately the main reason why unitaries live on a pedestal.

Eq. (2.24) doesn't really look like the Schrödinger equation you know. We can get to that by assuming we do a tiny evolution, from t to $t+\Delta t$. The operator U must of course satisfy $U(t,t)=1$ since this means we haven't evolved at all. Thus we can expand it in a Taylor series in Δt, which to first order can be written as

$$U(t+\Delta t,t) \simeq 1 - i\Delta t H(t) \tag{2.25}$$

where $H(t)$ is some operator which, as you of course know, is called the **Hamiltonian** of your system. The reason why I put an i in front is to make H Hermitian. I also didn't introduce Planck's constant $\hbar$. In this course $\hbar = 1$. This simply means that time and energy have the same units:

$$\boxed{\text{In this course we always set } \hbar = 1}$$

Inserting Eq. (2.25) in Eq. (2.24), dividing by Δt and then taking the limit $\Delta t \to 0$ we get

$$\boxed{\partial_t|\psi(t)\rangle = -iH(t)|\psi(t)\rangle} \tag{2.26}$$

which is Schrödinger's equation in it's more familiar form.

What we have learned so far is that, once we postulate normalization and linearity, the evolution of a physical system *must* be given by an equation of the form (2.26), where $H(t)$ is some operator. Thus, the structure of Schrödinger's equation is really a consequence of these two postulates. Of course, the really hard question is what is the operator $H(t)$. The answer is usually a mixture of physical principles and experimental observations.

If we plug Eq. (2.24) into Eq. (2.26) we get

$$\partial_t U(t, t_0)|\psi(t_0)\rangle = -iH(t)U(t, t_0)|\psi(t_0)\rangle.$$

Notice that this is an equation for t: the parameter t_0 doesn't do anything. Since the equation must hold for any initial state $|\psi(t_0)\rangle$, we can instead write down an equation for the unitary operator $U(t, t_0)$ itself. Namely,

$$\boxed{\partial_t U(t, t_0) = -iH(t)U(t, t_0), \qquad U(t_0, t_0) = 1.} \tag{2.27}$$

The initial condition here simply means that U must act trivially when $t = t_0$, since then there is no evolution to occur in the first place. Eq. (2.27) is also Schrödinger's equation, but written at the level of the time evolution operator itself. In fact, by linearity of matrix multiplications and the fact that $U(t_0, t_0) = 1$, each column of $U(t, t_0)$ can be viewed as the solution of Eq. (2.26) for an initial ket $|\psi_0\rangle = |i\rangle$. So Eq. (2.27) is essentially the same as solving Eq. (2.26) d times.

If the Hamiltonian is **time-independent**, then the solution of Eq. (2.27) is given by the time-evolution operator

$$U(t, t_0) = e^{-i(t-t_0)H}. \tag{2.28}$$

For time-dependent Hamiltonians one may also write a similar equation, but it will involve the notion of *time-ordering operator*. Defining the eigenstuff of the Hamiltonian as

$$H = \sum_n E_n|n\rangle\langle n|, \tag{2.29}$$

and using the tricks of Sec. 1.9, we may also write the evolution operator as

$$U(t, t_0) = \sum_n e^{-iE_n(t-t')}|n\rangle\langle n|. \tag{2.30}$$

Decomposing an arbitrary initial state as $|\psi_0\rangle = \sum_n \psi_n|n\rangle$ then yields

$$|\psi(t)\rangle = \sum_n e^{-iE_n(t-t_0)}\psi_n|n\rangle \tag{2.31}$$

Each component in the eigenbasis of the Hamiltonian evolves according to a simple exponential. Consequently, if the system starts in an eigenstate of the Hamiltonian, it stays there forever. On the other hand, if the system starts in a state which is not an eigenstate, it will simply continue to oscillate back and forth.

The von Neumann equation

We now translate Eq. (2.24) from kets to density matrices. Let us consider again the preparation idea that led us to Eq. (2.3). We have a machine that prepares states $|\psi_i(t_0)\rangle$ with probability p_i, so that the initial density matrix of the system is

$$\rho(t_0) = \sum_i p_i |\psi_i(t_0)\rangle\langle\psi_i(t_0)|.$$

These states then start to evolve in time. Each $|\psi_i(t_0)\rangle$ will therefore transform to $|\psi_i(t)\rangle = U(t,t_0)|\psi_i(t_0)\rangle$. Consequently, the evolution of $\rho(t)$ will be given by

$$\rho(t) = \sum_i p_i U(t,t_0)|\psi_i(t_0)\rangle\langle\psi_i(t_0)|U^\dagger(t,t_0).$$

But notice how we can factor out the unitaries $U(t,t_0)$ since they are the same for each ψ_i. As a consequence, we find

$$\rho(t) = U(t,t_0)\rho(t_0)U^\dagger(t,t_0).$$

To evolve a state, simply sandwich it on the unitary.

Differentiating with respect to t and using Eq. (2.27) also yields

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -iH(t)U(t,t_0)\rho(t_0)U^\dagger(t,t_0) + U(t,t_0)\rho(t_0)U^\dagger(t,t_0)iH(t) = -iH(t)\rho(t) + i\rho(t)H(t)$$

This gives us the differential form of the same equation. Let me summarize everything into one big box:

$$\boxed{\frac{\mathrm{d}\rho}{\mathrm{d}t} = -i[H(t),\rho], \qquad \rho(t) = U(t,t_0)\rho(0)U(t,t_0),} \tag{2.32}$$

which is von Neumann's equation.

An interesting consequence of Eq. (2.32) is that the purity (2.13) remains constant during unitary evolutions. To see that, we only need to use the cyclic property of the trace:

$$\mathrm{tr}\left\{\rho(t)^2\right\} = \mathrm{tr}\left\{U\rho_0 U^\dagger U\rho_0 U^\dagger\right\} = \mathrm{tr}\left\{\rho_0\rho_0 U^\dagger U\right\} = \mathrm{tr}\left\{\rho_0^2\right\}.$$

Thus,

$$\boxed{\text{The purity is constant in unitary evolutions.}}$$

This makes some sense: unitaries are like rotations. And a rotation should not affect the radius of the state (in the language of Bloch's sphere). Notwithstanding, this fact has some deep consequences, as we will soon learn.

2.6 Quantum operations

Von Neumann's equation is nothing but a consequence of Schrödinger's equation. And Schrödinger's equation was *entirely* based on two principles: linearity and normalization. Remember: we constructed it by asking, what is the kind of equation which preserves normalization and is still linear. And *voilá*, we had the unitary.

But we only asked for normalization of *kets*. The unitary is the kind of operator that preserves $\langle\psi(t)|\psi(t)\rangle = 1$. However, we now have density matrices and we can simply ask the question once again: what is the *most general* kind of operation which preserves the normalization of *density matrices*. Remember, a physical density matrix is any operator which is positive semidefinite and has trace $\text{tr}(\rho) = 1$. One may then ask, are there other linear operations, besides the unitary, which map physical density matrices to physical density matrices. The answer, of course, is yes! They are called quantum operations or quantum channels. And they are beautiful. :)

We will discuss the formal theory of quantum operations later on, when we have more tools at our disposal. Here I just want to illustrate one result due to Kraus.[1] A quantum operation can always be implemented as

$$\boxed{\mathcal{E}(\rho) = \sum_k M_k \rho M_k^\dagger, \qquad \sum_k M_k^\dagger M_k = 1.} \tag{2.33}$$

Here $\{M_k\}$ is an arbitrary *set* of operators, called **Kraus operators**, which only need to satisfy this condition that they sum to the identity. The unitary evolution is a particular case in which we have only one operator in the set $\{M_k\}$. Then normalization implies $U^\dagger U = 1$. Operations of this form are called **Completely Positive Trace Preserving (CPTP)** maps. Any map of the form (2.33) with the $\{M_k\}$ satisfying $\sum_k M_k^\dagger M_k = 1$ is, in principle, a CPTP map.

Amplitude damping

The most famous, and widely used, quantum operation is the amplitude damping channel. It is defined by the following set of operators:

$$M_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix}, \qquad M_1 = \begin{pmatrix} 0 & \sqrt{\lambda} \\ 0 & 0 \end{pmatrix}, \tag{2.34}$$

with $\lambda \in [0,1]$. This is a valid set of Kraus operators since $M_0^\dagger M_0 + M_1^\dagger M_1 = 1$. Its action on a general qubit density matrix of the form (2.16) is:

$$\mathcal{E}(\rho) = \begin{pmatrix} \lambda + p(1-\lambda) & q\sqrt{1-\lambda} \\ q^* \sqrt{1-\lambda} & (1-\lambda)(1-p) \end{pmatrix}. \tag{2.35}$$

---

[1] c.f. K. Kraus, *States, Effects and Operations: Fundamental Notions of Quantum Theory*, Springer Verlag 1983.

If $\lambda = 0$ nothing happens, $\mathcal{E}(\rho) = \rho$. Conversely, if $\lambda = 1$ then

$$\mathcal{E}(\rho)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.36}$$

for any initial density matrix ρ. This is why this is called an **amplitude damping**: no matter where you start, the map tries to push the system towards $|0\rangle$. It does so by destroying coherences, $q \to q\sqrt{1-\lambda}$, and by affecting the populations, $p \to \lambda + p(1-\lambda)$. The larger the value of λ, the stronger is the effect.

2.7 Generalized measurements

We are now in the position to introduce generalized measurements in quantum mechanics. As with some of the other things discussed in this chapter, I will not try to demonstrate where these generalized measurements come from right now (we will try to do that later). Instead, I just want to give you a new postulate for measurements and then explore how it looks like for some simple examples. The measurement postulate can be formulated as follows.

Generalized measurement postulate: any quantum measurement is fully specified by a set of Kraus operators $\{M_k\}$ satisfying

$$\boxed{\sum_k M_k^\dagger M_k = 1.} \tag{2.37}$$

The values of k label the possible outcomes of an experiment, which will occur randomly with each repetition of the experiment. The probability of obtaining outcome k is

$$\boxed{p_k = \text{tr}(M_k \rho M_k^\dagger).} \tag{2.38}$$

If the outcome of the measurement is k, then the state after the measurement will be

$$\boxed{\rho \to \frac{M_k \rho M_k^\dagger}{p_k},} \tag{2.39}$$

which describes the **measurement backaction** on the system.

The case of projective measurements, studied in Sec. 1.7, simply corresponds to taking $M_k = |k\rangle\langle k|$ to be projection operators onto some basis $|k\rangle$. Eqs. (2.37)-(2.39) then reduce to

$$\sum_k |k\rangle\langle k| = 1, \qquad p_k = \text{tr}(|k\rangle\langle k\rho) = \langle k|\rho|k\rangle, \qquad \rho \to \frac{P_k \rho P_k}{p_k} = |k\rangle\langle k|, \tag{2.40}$$

which are the usual projective measurement/full collapse scenario.

If the state happens to be pure, $\rho = |\psi\rangle\langle\psi|$, then Eqs. (2.38) and (2.39) also simplify slightly to

$$p_k = \langle\psi|M_k^\dagger M_k|\psi\rangle, \qquad |\psi\rangle \to \frac{M_k|\psi\rangle}{\sqrt{p_k}}. \tag{2.41}$$

The final state after the measurement continues to be a pure state. Remember that pure states contain no classical uncertainty. Hence, this means that performing a measurement does not introduce any uncertainty.

Suppose, however, that one performs the measurement but does not check the outcome. There was a measurement backaction, we just don't know exactly which backaction happened. So the best guess we can make to the new state of the system is just a statistical mixture of the possible outcomes in Eq. (2.39), each weighted with probability p_k. This gives

$$\rho' = \sum_k p_k \frac{M_k \rho M_k^\dagger}{p_k} = \sum_k M_k \rho M_k^\dagger, \tag{2.42}$$

which is nothing but the quantum operations introduced in Eq. (2.33). This gives a neat interpretation of quantum operations: they are statistical mixtures of making measurements but not knowing what the measurement outcomes were. This is, however, only *one* of multiple interpretations of quantum operations. Eq. (2.33) is just mathematics: how to write a CPTP map. Eq. (2.42), on the other hand, is physics.

Example: Amplitude damping

As an example, consider the amplitude damping Kraus operators in Eq. (2.34) and suppose that initially the system was prepared in the state $|+\rangle$. The probabilities of the two outcomes are

$$p_0 = 1 - \frac{\lambda}{2}, \tag{2.43}$$

$$p_1 = \frac{\lambda}{2}. \tag{2.44}$$

So in the case of a trivial channel ($\lambda = 0$) we get $p_1 = 0$. Conversely, in the case $\lambda = 1$ (full damping) we get $p_0 = p_1 = 1/2$. The backaction on the state of the system given that each outcome happened will be

$$|+\rangle \to \frac{M_0|+\rangle}{\sqrt{p_0}} = \frac{|0\rangle + \sqrt{1-\lambda}|1\rangle}{\sqrt{2-\lambda}}, \tag{2.45}$$

$$|+\rangle \to \frac{M_1|+\rangle}{\sqrt{p_1}} = |0\rangle. \tag{2.46}$$

The physics of outcome $k = 0$ seems a bit strange. But that of $k = 1$ is quite clear: if that outcome happened it means the state collapsed to $|0\rangle$.

This kind of channel is used to model photon emission. If you imagine the qubit is an atom and $|0\rangle$ is the ground state, then the outcome $k = 1$ means the atom emitted a photon. If that happened, then no matter which state it was, it will collapse to $|0\rangle$. The outcome $k = 0$ can then be interpreted as *not emitting a photon*. What is cool is that this doesn't mean nothing happened to the state. Even though it doesn't emit a photon, the state was still updated.

POVMs

Let us now go back to the probability of the outcomes, Eq. (2.38). Notice that as far as the probabilities are concerned, we don't really need to know the measurement operators M_k. It suffices to know

$$\boxed{E_k = M_k^\dagger M_k,} \tag{2.47}$$

which then gives

$$\boxed{p_k = \mathrm{tr}(E_k \rho).} \tag{2.48}$$

The set of operators $\{E_k\}$ are called a **Positive Operator Value Measure (POVM)**. By construction, the E_k are positive semidefinite operators. Moreover, the normalization (2.37) is translated to

$$\boxed{\sum_k E_k = 1.} \tag{2.49}$$

Any set of operators $\{E_k\}$ which are positive semidefinite and satisfy this normalization condition represents a valid POVM.

It is interesting to differentiate between generalized measurements and POVMs because different sets $\{M_k\}$ can actually produce the same set $\{E_k\}$. For instance, in the amplitude damping example, instead of Eq. (2.34), we can choose Kraus operators

$$M_0' = M_0, \qquad M_1' = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\lambda} \end{pmatrix}, \tag{2.50}$$

The backaction caused by M_1' will be different than that caused by M_1. However, one may readily check that

$$M_1^\dagger M_1 = (M_1')^\dagger (M_1') = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}.$$

Thus, even though the backaction is different, the POVM may be the same. This means that, as far as the probability of outcomes is concerned, both give exactly the same values.

In many experiments one does not have access to the *post-measurement state*, but only to the probability outcomes. In this case, all one needs to talk about are POVMs. We can also look at this from a more mathematical angle: what is the most general kind of mapping which takes as input a density matrix ρ and outputs a set of probabilities $\{p_k\}$? The answer is a POVM (2.48), with positive semidefinite E_k satisfying (2.49).

Informationally complete POVMs

A set of POVM elements $\{E_k\}$ yields a set of probabilities p_k. If we can somehow reconstruct the entire density matrix ρ from the numbers p_k, we say the POVM set $\{E_k\}$ is **informationally complete**. The density matrix of a d-dimensional system has $d^2 - 1$ independent elements (the -1 comes from normalization). POVMs are also subject to the normalization $\sum_k E_k = 1$. Thus, an informationally complete POVM must have at least d^2 elements.

For a qubit, this therefore means we need 4 elements at least. A symmetric example is the tetrahedron formed by the following projection operators $E_k = \frac{1}{2}|\psi_k\rangle\langle\psi_k|$, with

$$\begin{aligned} |\psi_1\rangle &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad & |\psi_2\rangle &= \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ \sqrt{2} \end{pmatrix}, \\ |\psi_3\rangle &= \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ \lambda\sqrt{2} \end{pmatrix}, \qquad & |\psi_2\rangle &= \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ \lambda^*\sqrt{2} \end{pmatrix}, \end{aligned} \tag{2.51}$$

where $\lambda = e^{2i\pi/3}$. These states are illustrated in Fig. 2.2.

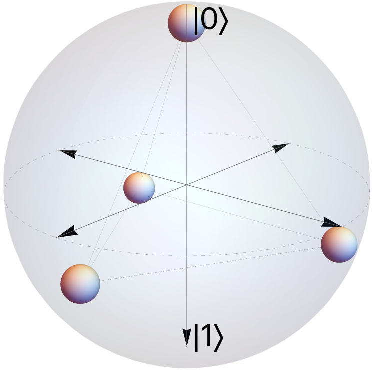

Figure 2.2: The tetrahedral informationally complete POVM for a qubit.

Another example, which the experimentalists prefer due to its simplicity, is the set of 6 Pauli states, $E_k = \frac{1}{\sqrt{6}}|\psi_k\rangle\langle\psi_k|$, with

$$\begin{aligned}
|\psi_1\rangle &= |z_+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, & |\psi_2\rangle &= |z_-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\
|\psi_3\rangle &= |x_+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, & |\psi_4\rangle &= |x_-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \\
|\psi_5\rangle &= |y_+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, & |\psi_6\rangle &= |y_-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}.
\end{aligned} \tag{2.52}$$

This POVM is not minimal: we have more elements than we need in principle. But from an experimental point of view that is actually a good thing, as it means more data is available.

2.8 The von Neumann Entropy

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, quite like the purity $\text{tr}(\rho^2)$. 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". We will also introduce another extremely important concept, called **relative entropy** which plays the role of a "distance" between two density matrices.

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

$$\boxed{S(\rho) = -\,\text{tr}(\rho \log \rho) = -\sum_k \lambda_k \log \lambda_k,} \tag{2.53}$$

where λ_k are the eigenvalues of ρ. Working with the logarithm of an operator can be awkward. That is why in the last equality I expressed $S(\rho)$ in terms of them. The logarithm in Eq. (2.53) can be either base 2 or base e. It depends if the application is more oriented towards information theory or physics (respectively). The last expression in (2.53), in terms of a sum of probabilities, is also called the **Shannon entropy**.

The entropy is seen to be a sum of functions of the form $-p\log(p)$, where $p \in [0,1]$. The behavior of this function is shown in Fig. 2.3. It tends to zero both when $p \to 0$ and $p \to 1$, and has a maximum at $p = 1/e$. Hence, any state which has $p_k = 0$ *or* $p_k = 1$ will not contribute to the entropy (even though $\log(0)$ alone diverges, $0\log(0)$ is well behaved). States that are too deterministic therefore contribute little to the entropy. Entropy likes randomness.

Since each $-p\log(p)$ is always non-negative, the same must be true for $S(\rho)$:

$$S(\rho) \geq 0. \tag{2.54}$$

Moreover, if the system is in a pure state, $\rho = |\psi\rangle\langle\psi|$, then it will have one eigenvalue $p_1 = 1$ and all others zero. Consequently, in a pure state the entropy will be zero:

$$\boxed{\text{The entropy of a pure state is zero.}} \tag{2.55}$$

In information theory the quantity $-\log(p_k)$ is sometimes called the *surprise*. When an "event" is rare ($p_k \sim 0$) this quantity is big ("surprise!") and when an event is common ($p_k \sim 1$) this quantity is small ("meh"). The entropy is then interpreted as the *average surprise* of the system, which I think is a little bit funny.

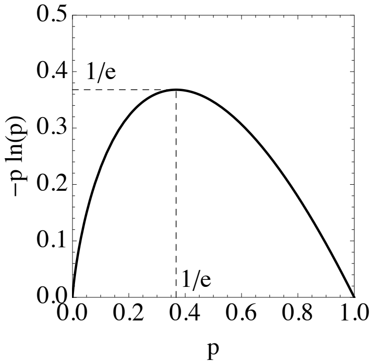

Figure 2.3: The function $-p\log(p)$, corresponding to each term in the von Neumann entropy (2.53).

As we have just seen, the entropy is bounded from below by 0. But if the Hilbert space dimension d is finite, then the entropy will also be bounded from above. I will leave this proof for you as an exercise. What you need to do is maximize Eq. (2.53) with respect to the p_k, but using Lagrange multipliers to impose the constraint $\sum_k p_k = 1$. Or, if you are not in the mood for Lagrange multipliers, wait until Eq. (**??**) where I will introduce a much easier method to demonstrate the same thing. In any case, the result is

$$\max(S) = \log(d). \qquad \text{Occurs when } \rho = \frac{\mathbb{I}}{d}. \tag{2.56}$$

The entropy therefore varies between 0 for pure states and $\log(d)$ for maximally disordered states. Hence, it clearly serves as a measure of how mixed a state is.

Another very important property of the entropy (2.53) is that it is invariant under unitary transformations:

$$S(U\rho U^\dagger) = S(\rho). \tag{2.57}$$

This is a consequence of the infiltration property of the unitaries $Uf(A)U^\dagger = f(UAU^\dagger)$ [Eq. (1.58)], together with the cyclic property of the trace. Since the time evolution of closed systems are implemented by unitary transformations, this means that the entropy is a constant of motion. We have seen that the same is true for the purity: unitary evolutions do not change the mixedness of a state. Or, in the Bloch sphere picture, unitary evolutions keep the state on the same spherical shell. For open quantum systems this will no longer be the case.

As a quick example, let us write down the formula for the entropy of a qubit. Recall the discussion in Sec. 2.4: the density matrix of a qubit may always be written as in Eq. (2.19). The eigenvalues of ρ are therefore $(1 \pm s)/2$ where $s = \sqrt{s_x^2 + s_y^2 + s_z^2}$ represents the radius of the state in Bloch's sphere. Hence, applying Eq. (2.53) we get

$$S = -\left(\frac{1+s}{2}\right)\log\left(\frac{1+s}{2}\right) - \left(\frac{1-s}{2}\right)\log\left(\frac{1-s}{2}\right). \tag{2.58}$$

For a pure state we have $s = 1$ which then gives $S = 0$. On the other hand, for a maximally disordered state we have $s = 0$ which gives the maximum value $S = \log 2$, the log of the dimension of the Hilbert space. The shape of S is shown in Fig. 2.4.

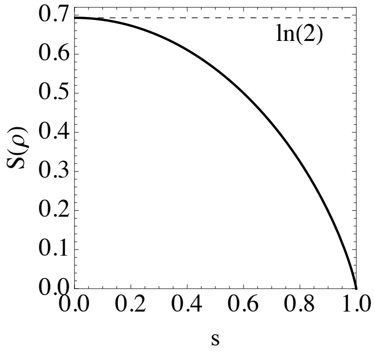

Figure 2.4: The von Neumann entropy for a qubit, Eq. (2.58), as a function of the Bloch-sphere radius s.

The quantum relative entropy

Another very important quantity in quantum information theory is the *quantum relative entropy* or *Kullback-Leibler divergence*. Given two density matrices ρ and σ, it is defined as

$$\boxed{S(\rho||\sigma) = \text{tr}(\rho\log\rho - \rho\log\sigma).} \tag{2.59}$$

This quantity is important for a series of reasons. But one in particular is that it satisfies the *Klein inequality*:

$$\boxed{S(\rho||\sigma) \geq 0, \qquad S(\rho||\sigma) = 0 \text{ iff } \rho = \sigma.} \tag{2.60}$$

The proof of this inequality is really boring and I'm not gonna do it here. You can find it in Nielsen and Chuang or even in Wikipedia.

Eq. (2.60) gives us the idea that we could use the relative entropy as a measure of the *distance* between two density matrices. But that is not entirely precise since the relative entropy does not satisfy the triangle inequality

$$d(x,z) \leq d(x,y) + +d(y,z). \tag{2.61}$$

This is something a true measure of distance must always satisfy. If you are wondering what quantities are actual distances, the *trace distance* is one of them[2]

$$\mathcal{T}(\rho,\sigma) = \|\rho-\sigma\|_1 := \operatorname{tr}\left[\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}\right]. \tag{2.62}$$

But there are others as well.

Entropy and information

Define the maximally mixed state $\pi = I_d/d$. This is the state we know absolutely nothing about. We have zero information about it. Motivated by this, we can define the amount of information in a state ρ as the "distance" between ρ and π; viz,

$$\mathcal{I}(\rho) = S(\rho\|\pi).$$

But we can also open this up as

$$S(\rho\|1/d) = \operatorname{tr}(\rho\log\rho) - \operatorname{tr}(\rho\log(I_d/d)) = -S(\rho) + \log(d).$$

I know it is a bit weird to manipulate I_d/d here. But remember that the identity matrix satisfies *exactly* the same properties as the number one, so we can just use the usual algebra of logarithms in this case.

We see from this result that the information contained in a state is nothing but

$$\boxed{\mathcal{I}(\rho) = S(\rho\|\pi) = \log(d) - S(\rho).} \tag{2.63}$$

This shows how information is connected with entropy. The larger the entropy, the less information we have about the system. For the maximally mixed state $S(\rho) = \log(d)$ and we get zero information. For a pure state $S(\rho) = 0$ and we have the maximal information $\log(d)$.

As I mentioned above, the relative entropy is very useful in proving some mathematical relations. For instance consider the result in Eq. (2.56). If we look at Eq. (2.63) and remember that $S(\rho\|\sigma) \geq 0$, this result becomes kind of obvious: $S(\rho) \leq \log(d)$ and $S(\rho) = \log(d)$ iff $\rho = 1/d$, which is precisely Eq. (2.56).

---

[2] The fact that $\rho - \sigma$ is Hermitian can be used to simplify this a bit. I just wanted to write it in a more general way, which also holds for non-Hermitian operators.

Chapter 3

Composite Systems

3.1 The age of Ulkron

So far we have considered only a single quantum system described by a basis $|i\rangle$. Now we turn to the question of how to describe mathematically a system composed of two or more sub-systems. Suppose we have two sub-systems, which we call A and B. They can be, for instance, two qubits: one on earth and the other on mars. How to write states and operators for this joint system? This is another postulate of quantum theory. But instead of postulating it from the start, I propose we first try to formulate what we intuitively expect to happen. Then we introduce the mathematical framework that does the job.

For me, at least, I would expect the following to be true. First, if $\{|i\rangle_A\}]$ is a set of basis vectors for A and $\{|j\rangle_B\}$ is a basis vector for B, then a joint basis for AB should have the form $|i, j\rangle$. For instance, for two qubits one should have four possibilities:

$$|0,0\rangle, \qquad |0,1\rangle, \qquad |1,0\rangle, \qquad |1,1\rangle.$$

Secondly, again at least in my intuition, one should be able to write down operators that act *locally* as if the other system was not there. For instance, we know that for a single qubit σ_x is the bit flip operator:

$$\sigma_x|0\rangle = |1\rangle, \qquad \sigma_x|1\rangle = |0\rangle.$$

If we have two qubits, I would expect we should be able to define two operators σ_x^A and σ_x^B that act as follows:

$$\sigma_x^A|0,0\rangle = |1,0\rangle, \qquad \sigma_x^B|0,0\rangle = |0,1\rangle.$$

Makes sense, no? Similarly, we expect that if we apply both σ_x^A and σ_x^B the order shouldn't matter:

$$\sigma_x^A\sigma_x^B|0,0\rangle = \sigma_x^B\sigma_x^A|0,0\rangle = |1,1\rangle.$$

This means that operators belonging to different systems should commute:

$$[\sigma_x^A, \sigma_x^B] = 0. \tag{3.1}$$

The tensor/Kronecker product

The mathematical structure that implements these ideas is called the **tensor product** or **Kronecker product**. It is, in essence, a way to glue together two vector spaces to form a larger space. The tensor product between two states $|i\rangle_A$ and $|j\rangle_B$ is written as

$$|i, j\rangle = |i\rangle \otimes |j\rangle. \tag{3.2}$$

The symbol $\otimes$ separates the two universes. We read this as "i tens j" or "i kron j". I like the "kron" since it reminds me of a crappy villain from a Transformers or Marvel movie. Similarly, the operators σ_x^A and σ_x^B are defined as

$$\sigma_x^A = \sigma_x \otimes I, \qquad \sigma_x^B = I \times \sigma_x, \tag{3.3}$$

where I is the identify matrix.

In order for us to make sense of these definitions, we must of course specify the basic rules for how objects behave around the $\otimes$. Lucky for you, there is only one rule that we really need to remeber: *stuff to the left of $\otimes$ only interact with stuff to the left and stuff to the right only interact with stuff to the right*. In symbols:

$$\boxed{(A \otimes B)(C \otimes D) = (AC) \otimes (BD),} \tag{3.4}$$

In this rule A, B, C and D can be *any* mathematical object, as long as the multiplications AC and BD make sense.

Let's see how this works. For instance,

$$\sigma_x^A |0, 0\rangle = (\sigma_x \otimes I)(|0\rangle \otimes |0\rangle) = (\sigma_x|0\rangle) \otimes (I|0\rangle).$$

The only thing I did was apply the rule (3.4) to combine stuff to the left of $\otimes$ with stuff to the left and stuff to the right with stuff to the right. Now that we have $\sigma_x|0\rangle$ we are back to the single qubit business, so we can just write $\sigma_x|0\rangle = |1\rangle$. Then we recombine the result:

$$(\sigma_x|0\rangle) \otimes (I|0\rangle) = |1\rangle \otimes |0\rangle = |1, 0\rangle,$$

which is what we would expect intuitively. As another example, the property (3.1), that operators pertaining to different systems should commute, now follows directly from our definitions:

$$\sigma_x^A \sigma_x^B = (\sigma_x \otimes I)(I \otimes \sigma_x) = (\sigma_x \otimes \sigma_x),$$

$$\sigma_x^B \sigma_x^A = (I \otimes \sigma_x)(\sigma_x \otimes I) = (\sigma_x \otimes \sigma_x),$$

which are definitely the same thing.

Everything we just said also holds for systems composed of 3, 4 or any number of parts, of course. In this case we simply add more and more $\otimes$. For instance, for 3 qubits, $\sigma_x^B = I \otimes \sigma_x \otimes I$ and so on.

Let us also talk about how to combine other kinds of objects. Remember that all we need is for the multiplications in the composition rule (3.4) to make sense. For instance, an operation that makes no sense is

$$(\langle 0| \otimes |0\rangle)(\sigma_x \otimes \sigma_x) = \text{crazy nonsense},$$

because even though $\langle 0|\sigma_x$ makes sense, the operation $|0\rangle\sigma_x$ does not.

An operation which does make sense is

$$\langle k, \ell|i, j\rangle = (\langle k| \otimes \langle \ell|)(|i\rangle \otimes |j\rangle) = (\langle k|i\rangle) \otimes (\langle \ell|j\rangle).$$

The objects that remain here are two numbers and the tensor product of two numbers is also a number. Thus, we arrive at a rule for the inner product:

$$\langle k, \ell|i, j\rangle = \langle k|i\rangle\langle \ell|j\rangle. \tag{3.5}$$

Outer products are similarly defined:

$$|k, \ell\rangle\langle i, j| = |k\rangle\langle i| \otimes |\ell\rangle\langle j|. \tag{3.6}$$

One can also come up with somewhat weird operations which nonetheless make sense. For instance,

$$(\langle k| \otimes |\ell\rangle)(|i\rangle \otimes \langle j|) = (\langle k|i\rangle) \otimes |\ell\rangle\langle j| = (\langle k|i\rangle)|\ell\rangle\langle j|.$$

In the last equality I used the fact that $\langle k|i\rangle$ is just a number.

Be cool about notation

Here is a really really really good tip: *be cool duuuuuude*. There are many ways of expressing quantum states and operators for composite systems. *Don't be rigid about notation. Just be clear so that people know what you mean.* For instance, if we talk about states, the following notations are equivalent:

$$\boxed{|i, j\rangle_{AB} = |i\rangle_A \otimes |j\rangle_B = |i\rangle_A|j\rangle_B.} \tag{3.7}$$

In the third notation adding the suffixes A and B is essential. Otherwise one would not know if $|i\rangle$ belongs to A or B. For completeness I also added the suffixes to the first two notations. Sometimes that is redundant. But if there is ever room for confusion, add it: it doesn't cost much.

A notation like $|i\rangle_A|j\rangle_B$ also allows you to move things around and write, for instance, $|j\rangle_B|i\rangle_A$. There is no room for confusion because you know one symbol belongs to A and the other to B. The same is true for operator multiplication. For instance,

$$\sigma_x^B|i, j\rangle_{AB} = |i\rangle_A\sigma_x^B|j\rangle_B.$$

Notice that there is zero room for misinterpretation: the notation is not rigid, but no one will interpret it wrong.

I *strongly* recommend you be cool about the notation. Each notation is useful for a different thing, so feel free to change them at will. Just make sure there is no room for misinterpretation.

Matrix representation of the Kronecker product

When using the Kronecker product in a computer, it is standard to order the basis elements $|i, j\rangle$ in *lexicographic order*: for each entry of the first, you loop over all elements of the last. For instance, if A and B have each dimension 3, we get

$$|0,0\rangle, \quad |0,1\rangle, \quad |0,2\rangle, \quad |1,0\rangle, \quad |1,1\rangle, \quad |1,2\rangle, \quad |2,0\rangle, \quad |2,1\rangle, \quad |2,2\rangle.$$

Conversely, if we have 3 qubits, we would order the basis elements as

$$|0,0,0\rangle, \qquad |0,0,1\rangle, \qquad |0,1,0\rangle, \qquad |0,1,1\rangle \qquad |1,0,0\rangle \quad \dots$$

This ordering is not mandatory. But is extremely convenient for the following reason.

We then associate to each element a unit vector. For instance, for 2 qubits we would have,

$$|0,0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad |0,1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad |1,0\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad |1,1\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{3.8}$$

The matrix elements of an operator of the form $A \times B$ then becomes, using the property (3.4)

$$\langle k, \ell | A \otimes B | i, j \rangle = \langle k | A | i \rangle \langle \ell | B | j \rangle = A_{ki} B_{\ell j}.$$

If we now present these guys in a matrix, since we loop over all elements of the second index, for each element of the first, the matrix form of this will look like

$$A \otimes B = \begin{pmatrix} A_{0,0} B & \dots & a_{0,d_A-1} B \\ \vdots & \ddots & \vdots \\ a_{d_A-1,0} B & \dots & a_{d_A-1,d_A-1} B \end{pmatrix}. \tag{3.9}$$

This is just an easy of visualizing the matrix: for each A_{ki} we introduce a full block B. To be clear what is meant by this, consider for instance

$$\sigma_x \otimes \sigma_x = \begin{pmatrix} 0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ 1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.10}$$

This provides an automated way to construct tensor product matrices. The final result is not very intuitive. But computationally, it is quite trivial. Specially since the Kronecker product is implemented in any library. In MATLAB they call it `kron()` whereas in Mathematica they call it `KroneckerProduct[]`. These functions are really useful. You should really try to play with them a bit.

As a consistency check, we can verify that the same logic also holds for vectors. For instance,

$$|0,0\rangle = |0\rangle \otimes |0\rangle = \begin{pmatrix} 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ 0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{3.11}$$

Proceeding similarly leads to all elements in Eq. (3.8).

3.2 Entanglement and mixed states

So far we have talked about how to represent kets and operators of composite systems using the tensor product. Now let's see what kind of physics this produces. Suppose we have two qubits, A and B. If qubit A is on Earth and qubit B is on Mars, it is reasonable to assume that they are each in *local states*, such as

$$|\eta\rangle_A = \alpha|0\rangle_A + \beta|1\rangle_A, \qquad |\phi\rangle_B = \gamma|0\rangle_B + \delta|1\rangle_B.$$

Then, the global state of AB will be

$$\begin{aligned}|\eta\rangle_A \otimes |\phi\rangle_B &= \Big[\alpha|0\rangle_A + \beta|1\rangle_A\Big] \otimes \Big[\gamma|0\rangle_B + \delta|1\rangle_B\Big] \\ &= \alpha\gamma|0,0\rangle_{AB} + \alpha\delta|0,1\rangle_{AB} + \beta\gamma|1,0\rangle_{AB} + \beta\delta|1,1\rangle_{AB}.\end{aligned}$$

If we look at the second line, this state seems like simply a linear combination of the four basis elements $|i, j\rangle_{AB}$. However, this is not an arbitrary linear combination. It contains a very special choice of parameters which are such that you can perfectly factor the state into something related to A times something related to B. Cases like this are what we call a **product state**. If A and B are in a product state, they are completely independent of each other.

However, quantum theory also allows us to have more general linear combinations which are not necessarily factorable into a product. Such a general linear combination has the form

$$|\psi\rangle_{AB} = \sum_{i,j} \psi_{ij}|i, j\rangle_{AB}, \tag{3.12}$$

where $\psi_{i,j}$ are any set of complex numbers satisfying $\sum_{ij} |\psi_{ij}|^2 = 1$. When a state like this cannot be written as a product,[1] we say A **and** B **are entangled**. An important set of entangled states are the so called **Bell states**:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\Big[|0,0\rangle + |1,1\rangle\Big], \tag{3.13}$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}\Big[|0,0\rangle - |1,1\rangle\Big], \tag{3.14}$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}\Big[|0,1\rangle + |1,0\rangle\Big], \tag{3.15}$$

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}\Big[|0,1\rangle - |1,0\rangle\Big]. \tag{3.16}$$

These states *cannot* be factored into a product of local states (please take a second to convince yourself of that!). In fact, we will learn soon that they are **maximally**

---

[1] That is, when we cannot decompose $\psi_{ij} = f_j g_j$.

entangled states. If you are familiar with the theory of angular momentum, you will also notice that these states (specially $|\Psi^\pm\rangle$) are exactly to the singlet and triplet states of two spin 1/2 particles. Moreover, it is useful to note that they form an orthonormal basis for the Hilbert space of the two qubits.

The Controlled NOT (CNOT)

We usually entangle systems by applying *gates*. That is, unitary transformations stemming from the interaction between the systems. A popular entangling gate for two qubits is the CNOT. It is defined by the unitary

$$U_{\text{CNOT}} = |0\rangle\langle 0|_A \otimes I_B + |1\rangle\langle 1|_A \otimes \sigma_x^B. \tag{3.17}$$

Qubit A is the *control bit*. If it is in $|0\rangle$, we do nothing on B. But if it is in $|1\rangle$, we apply the bit flip operation σ_x on B:

$$\begin{aligned} U_{\text{CNOT}}|0\rangle_A|\psi\rangle_B &= |0\rangle_A|\psi\rangle_B, \\ U_{\text{CNOT}}|0\rangle_A|\psi\rangle_B &= |0\rangle_A\Big(\sigma_x^B|\psi\rangle_B\Big). \end{aligned}$$

Suppose we now start with two qubits reset to $|0\rangle_A|0\rangle_B$. We can prepare the two qubits in a Bell state by applying two gates. First, we apply a Hadamard gate to A:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{3.18}$$

This produces

$$H_A|0\rangle_A|0\rangle_B = |+\rangle_A|0\rangle_B = \left(\frac{|0\rangle_A + |1\rangle_A}{\sqrt{2}}\right)|0\rangle_B.$$

This is a gate acting only on A. It is a *local operation* and thus cannot entangle A and B. To entangle them we now apply the CNOT (3.17). It gives

$$U_{\text{CNOT}}H_A|0\rangle_A|0\rangle_B = \frac{|0,0\rangle_{AB} + |1,1\rangle_{AB}}{\sqrt{2}},$$

which is nothing but the Bell state (3.13). The other Bell states may be generated in a similar way, by starting with the four possible states $|i, j\rangle$:

$$\begin{aligned} |\Phi^+\rangle &= U_{\text{CNOT}}H_A|0\rangle_A|0\rangle_B, \\ |\Phi^-\rangle &= U_{\text{CNOT}}H_A|1\rangle_A|0\rangle_B, \\ |\Psi^+\rangle &= U_{\text{CNOT}}H_A|0\rangle_A|1\rangle_B, \\ |\Psi^-\rangle &= U_{\text{CNOT}}H_A|1\rangle_A|1\rangle_B. \end{aligned} \tag{3.19}$$

Density matrices from entanglement

Now I want you to recall our original discussion in Sec. 2.1. We saw that the concept of density matrix naturally appeared when we considered a crappy machine that produced quantum states with some classical uncertainty. What we found was that it was possible to combine quantum and classical effects by introducing an object of the form

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{3.20}$$

where the $|\psi_i\rangle$ are arbitrary states and the p_i are arbitrary probabilities. This construction may have left you with the impression that the density matrix is only necessary when we want to mix quantum and classical stuff. That density matrices are not really a quantum thing. Now I want to show you that this is not the case. It is *definitely* not the case. I will show you that there is an intimate relation between mixed states and entanglement. And this relation is one the key steps relating quantum mechanics and information theory.

Essentially, the connection is made by the notion of **reduced state** or **reduced density matrix**. When a composite system is in a product state $|\psi\rangle_A \otimes |\phi\rangle_B$, it makes sense to say the state of A is simply $|\psi\rangle_A$. But if A and B are entangled, then what is exactly the "state" of A? To warm up, consider first a bipartite state of AB of the form

$$|\psi\rangle_{AB} = \sum_i c_i \; |i\rangle \otimes |i\rangle \tag{3.21}$$

for certain coefficients c_i satisfying $\sum_i |c_i|^2 = 1$. If $c_i = 1$ for some i and all other $c_j = 0$ then $|\psi\rangle = |i\rangle \otimes |i\rangle$ and we get a product state. In any other case, the state will be entangled.

Now let O_A be an operator which acts only on system A. That is, an operator which has the form $O_A = O_A \otimes I_B$. The expectation value of O_A in the state (3.21) will be

$$\langle O_A \rangle = \langle\psi|(O_A \otimes I_B)|\psi\rangle \tag{3.22}$$

Carrying out the calculation we get:

$$\begin{aligned}
\langle O_A \rangle &= \sum_{i,j} c_i^* c_j \; (\langle i| \otimes \langle i|)(O_A \otimes I_B)(|j\rangle \otimes |j\rangle) \\
&= \sum_{i,j} c_i^* c_j \; \langle i|O_A|j\rangle\langle i|j\rangle \\
&= \sum_i |c_i|^2 \langle i|O_A|i\rangle.
\end{aligned}$$

The sandwich that remains is now performed only over the reduced state of A. However, each sandwich $\langle i|O_A|i\rangle$ is now weighted by a factor $|c_i|^2$.

We now ask the following question: can we attribute a state $|\psi_A\rangle$ for system A such that the above result can be expressed as $\langle\psi_A|O_A|\psi_A\rangle$. This is actually the same question we asked in Sec. 2.1. And we saw that the answer is *no*. In general, there is no pure state we can associate with A. Instead, if we wish to associate a quantum state to A, it will have to be a mixed state, described by a density matrix of the form

$$\rho_A = \sum_i |c_i|^2 |i\rangle\langle i| \tag{3.23}$$

then the expectation value of A becomes

$$\langle A\rangle = \text{tr}(A\rho_A) \tag{3.24}$$

This result has *extremely* important consequences. Eq. (3.23) has exactly the same form as Eq. (3.20), with the classical probabilities p_i replaced by quantum coefficients $|c_i|^2$. But there is absolutely *nothing* classical here. Nothing. We started with a pure state. We are talking about a purely quantum effect. Notwithstanding, we see that in general the state of A will be mixed. If $c_i = 1$ for some i and all other $c_j = 0$ then Eq. (3.23) reduces to $\rho_A = |i\rangle\langle i|$, which is a pure state. In all other cases, the state of A will be mixed. Thus,

> When AB are entangled, the reduced state of A and B will be mixed.

To give an example, suppose AB is in the Bell state (3.13). This state has the form of Eq. (3.21) with $c_i = 1/\sqrt{2}$. Thus, it is easy to apply Eq. (3.23), which gives

$$\rho_A = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.25}$$

We therefore see that the reduced state of A is actually the maximally mixed state. This is a feature of all Bell states and it is the reason we call them maximally entangled states. This is super interesting, if you think about it: A Bell state is a pure state, so we know exactly what the state of AB is. However, we know absolutely *nothing* about A alone.

3.3 Reduced density matrices and the partial trace

The state ρ_A in Eq. (3.23) is called a **reduced density matrix**. And the procedure that led us from $|\psi\rangle_{AB}$ to ρ_A is called the partial trace. This is the quantum analog of computing the marginal $P(x)$ of a joint probability distribution $P(x, y)$. In this section I will teach you how to make this procedure in a more algorithmic way.

The partial trace

Consider a bipartite 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. Any operator in AB can always be decomposed as

$$O = \sum_\alpha A_\alpha \otimes B_\alpha, \tag{3.26}$$

for some index α and some set of operators A_α and B_α. So, to start, let us consider simply an operator of the form $O = A \otimes B$. Then, by linearity it will be easy to extend to more general operators.

We begin by computing the trace of $O = A \otimes B$ in the $|a, b\rangle$ basis:

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

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

$$\boxed{\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B).} \tag{3.27}$$

We started with an operator having support on two Hilbert spaces and ended up tracing everything, so that we are left with only a single number.

We can now imagine an operation where we only trace over one of the Hilbert spaces and obtain an operator still having support on the other part. This is what we call the **partial trace**. It is defined as

$$\boxed{\operatorname{tr}_A(A \otimes B) = \operatorname{tr}(A)B, \qquad \operatorname{tr}_B(A \otimes B) = A\operatorname{tr}(B)} \tag{3.28}$$

When you "trace over A", you eliminate the variables pertaining to A and what you are left with is an operator acting only on 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 O as defined in Eq. (3.26), we have

$$\operatorname{tr}_A O = \sum_\alpha \operatorname{tr}(A_\alpha) B_\alpha, \qquad \operatorname{tr}_B O = \sum_\alpha A_\alpha \operatorname{tr}(B_\alpha). \tag{3.29}$$

An important example is the partial trace of an operator of the form $|a, b\rangle\langle a', b'|$. 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.,

$$\operatorname{tr}_B |a, b\rangle\langle a', b'| = |a\rangle\langle a'| \operatorname{tr}\Big\{|b\rangle\langle b'|\Big\}$$

Thus, we conclude that

$$\boxed{\operatorname{tr}_A |a,b\rangle\langle a',b'| = \delta_{a,a'}\ |b\rangle\langle b'|, \qquad \operatorname{tr}_B |a,b\rangle\langle a',b'| = |a\rangle\langle a'|\ \delta_{b,b'}.} \tag{3.30}$$

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

$$\boxed{\rho_A = \operatorname{tr}_B \rho_{AB}, \qquad \rho_B = \operatorname{tr}_A \rho_{AB}} \tag{3.31}$$

Let us now practice with some examples.

Example: Bell states

To practice, consider the Bell state example that led us from the bipartite state (3.13) to the reduced state (3.25). The global density matrix is

$$\rho_{AB} = |\Phi_1\rangle\langle\Phi_1| = \frac{1}{2}\Big\{|0,0\rangle\langle 0,0| + |0,0\rangle\langle 1,1| + |1,1\rangle\langle 0,0| + |1,1\rangle\langle 1,1|\Big\} \tag{3.32}$$

To take the partial trace we use Eq. (3.30) to find:

$$\rho_A = \frac{1}{2}\Big\{|0\rangle\langle 0| + |1\rangle\langle 1|\Big\} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.33}$$

with an identical result for ρ_B.

Example: Partially entangled states

Consider now a state of the form

$$|\psi\rangle_{AB} = \sqrt{p}|0,1\rangle_{AB} + \sqrt{1-p}|0,1\rangle_{AB}, \tag{3.34}$$

for some number $p \in [0,1]$. If $p = 1/2$ we recover the Bell state (3.13). To take the partial trace we proceed as before:

$$\rho_{AB} = |\psi\rangle\langle\psi| = p|0,1\rangle\langle 0,1| + (1-p)|1,0\rangle\langle 1,0| + \sqrt{p(1-p)}\Big(|0,1\rangle\langle 1,0| + |1,0\rangle\langle 0,1|\Big).$$

Due to Eq. (3.30), the last two terms will always give 0 when we take the partial trace. We are then left with

$$\rho_A = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|,$$

$$\rho_B = (1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|.$$

Example: X states

X states of two qubits are density matrices of the form

$$\rho = \begin{pmatrix} p_1 & 0 & 0 & \beta \\ 0 & p_2 & \alpha & 0 \\ 0 & \alpha^* & p_3 & 0 \\ \beta^* & 0 & 0 & p_4 \end{pmatrix}. \tag{3.35}$$

Normalization requires $\sum_i p_i = 1$ and positivity imposes constraints on the allowed values of α and β. The ordering of states here is the lexicographic order we discussed in Sec. 3.1. Namely, $|0,0\rangle, |0,1\rangle, |1,0\rangle, |1,1\rangle$. Thus, we can write this X state more explicitly as

$$\begin{aligned} \rho &= p_1|0,0\rangle\langle 0,0| + p_2|0,1\rangle\langle 0,1| + p_3|1,0\rangle\langle 1,0| + p_4|1,1\rangle\langle 1,1| \\ &+\alpha|0,1\rangle\langle 1,0| + \alpha^*|1,0\rangle\langle 0,1| \\ &+\beta|0,0\rangle\langle 1,1| + \beta^*|1,1\rangle\langle 0,0|. \end{aligned} \tag{3.36}$$

The meaning of α and β now become a bit more clear: they represent, respectively, the non-local coherences between $\{|0,1\rangle, |1,0\rangle\}$ and $\{|0,0\rangle, |1,1\rangle\}$. From Eq. (3.36) it is easy to take the partial trace:

$$\rho_A = \operatorname{tr}_B \rho = \begin{pmatrix} p_1 + p_2 & 0 \\ 0 & p_3 + p_4 \end{pmatrix}, \tag{3.37}$$

$$\rho_B = \operatorname{tr}_A \rho = \begin{pmatrix} p_1 + p_3 & 0 \\ 0 & p_2 + p_4 \end{pmatrix}. \tag{3.38}$$

We see that for X states, the reduced density matrices are diagonal. The entries which are set to zero in Eq. (3.35) are precisely the ones that would lead to non-zero diagonals in the reduced state. If we now look for observables, for instance, we will then find that

$$\langle \sigma_x^A \rangle = \langle \sigma_x^B \rangle = \langle \sigma_y^A \rangle = \langle \sigma_y^B \rangle = 0.$$

Non-local observables, on the other hand, can be non-zero. For instance, one may check that

$$\langle \sigma_x^A \sigma_x^B \rangle = \alpha + \alpha^* + \beta + \beta^*.$$

Partial trace looses information

If we have a state which is of the form $\rho_{AB} = \rho_A \otimes \rho_B$, then Eq. (3.28) 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 (3.31), we have

$$\rho_A \otimes \rho_B \neq \rho_{AB} \tag{3.39}$$

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

Example: separable states

A state is called separable when it can be written in the form

$$\rho = \sum_i p_i \, \rho_A^i \otimes \rho_B^i, \tag{3.40}$$

for a set of probabilities $p_i \in [0,1]$, $\sum_i p_i = 1$ and an arbitrary set of density matrices $\rho_{A(B)}^i$. Of course, in light of Eq. (3.26), any density matrix of AB can be decomposed as a sum of products. But usually each term in the sum is not a valid density matrix with a valid probability. The reason why a state of the form (3.40) is physically interesting is because it represents a classical statistical mixture of states of A and B.

This is just like the crappy machine of Sec. 2.1. With some probability p_i the machine prepares a state ρ_A^i for A and ρ_B^i for B. The two systems will in general be correlated: if we learn something about A, we can usually infer something about B. But this is only because they share a **common ignorance** about the machine that produced them. The states of A and B may very well be quantum. But their correlations are purely classical. We say *separable states of the form (3.40) are not entangled.*

Classical states and Quantum Discord

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

$$\rho_{AB} = \sum_{i,j} p_{i,j} \, |ij\rangle\langle ij| \tag{3.41}$$

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 nothing but a *classical* probability distribution encoded in a density matrix. To compute the partial trace over B we use Eq. (3.30), which gives

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

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

$$p_i^A = \sum_j p_{i,j}. \tag{3.42}$$

This is precisely the marginalization procedure in classical probability theory. We simply sum over all probabilities of B to obtain a reduced probability distribution only for A. A state which is not of the form (3.41) is said to have some *quantum discord*. We will talk more about what this means soon, when we talk about measurements.

3.4 Measurements in bipartite systems

The measurement postulates discussed in Sec. 2.7 remain exactly the same for composite systems. In fact, what is neat about those postulates is precisely their flexibility: any set of operators $\{M_k\}$ satisfying $\sum_k M_k^\dagger M_k = 1$ constitutes a possible set of generalized measurement operators. The interesting new thing that appears when talking about composite systems is that now we can specify measurement operators that act only on one of the subsystems and then understand what is the measurement backaction on the other. Or we can also define non-local measurements that act on both systems simultaneously.

Example: maximally entangled states

To warm up with an example, suppose that Alice and Bob have two qubits prepared in a Bell state $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ [Eq. (3.13)]. Alice then performs a projective measurement on the computational basis. This will be described by measurement operators

$$P_i = |i\rangle\langle i|_A \otimes I_B, \qquad i \in \{0, 1\}, \tag{3.43}$$

meaning we project onto $|0\rangle$ and $|1\rangle$ on Alice's side and do nothing on Bob's. Applying the P_i onto the Bell state $|\Phi^+\rangle$ yields $P_0|\Phi^+\rangle = |00\rangle/\sqrt{2}$ and $P_1|\Phi^+\rangle = |11\rangle/\sqrt{2}$. The probability of the two outcomes will be, according to Eq. (2.41), $p_i = \langle\Phi^+|P_i|\Phi^+\rangle$ which, in this case, reduces simply to $p_0 = p_1 = 1/2$. Alice therefore gets 0 or 1 with equal probability. If Alice happens to get 0, however, the global state of Alice and Bob will collapse to

$$|\Phi^+\rangle \to \frac{P_0|\Phi^+\rangle}{\sqrt{p_0}} = |00\rangle. \tag{3.44}$$

And similarly, if Alice gets a 1, the state will collapse to $|11\rangle$.

The state of Bob before the measurement could have been either 0 or 1. But now, if Alice happened to get 0, the state of Bob collapsed to 0. The backaction of the measurement on Alice caused the state of Bob to collapse. The global state after the measurement will thus be either $|00\rangle$ or $|11\rangle$, each occurring with probability 1/2. If Alice does not happen to read the outcome of her measurement, then the best possible guess for the state will be the statistical mixture

$$\rho' = \frac{|00\rangle\langle 00| + |11\rangle\langle 11|}{2}. \tag{3.45}$$

This state is now a classical probabilistic mixture of the two possibilities. It is therefore fundamentally different from the original entangled state $|\Phi^+\rangle$.

No-signaling theorem

According to the previous example, if Alice performs a measurement on her qubit, it will instantaneously affect the state of Bob, irrespective of how far they are from each other. This is the spooky part of quantum theory. It suggests that this kind of process could be used to signal faster than the speed of light, thus violating causality. Although spooky, this result is correct and can be verified in the laboratory, as we will discuss in the next section. However, it cannot be used to transmit information between distant observers. That is the idea of the **no-signaling** (or no-communications) principle.

The reason why this is so is because as far as Bob is concerned, he cannot tell that Alice made the measurement. Before the measurement was made, Bob's state was a maximally mixed state

$$\rho_B = \text{tr}_A\,|\Phi^+\rangle\langle\Phi^+| = \frac{I_B}{2}.$$

When Alice performs her measurement, the global state will collapse to either 0 or 1, each with probability $1/2$. But there exist no measurement scheme which Bob can do to figure out which outcome she got. From a probabilistic point of view, the best guess Bob can make about his state is

$$\rho_B' = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{I_B}{2}.$$

The only way Bob can tell that a measurement was actually done would be to *classically communicate* with Alice (e.g. via Whatsapp). If Alice tell him that she obtained 0 in her measurement, then Bob would be able to update his state. Otherwise, he has no way to tell. Thus, local measurements cannot be used to signal information.

We can make this proof more formal, as follows. Consider a local measurement on Alice's side, described by a set of Kraus operators $\{M_k^A = M_k \otimes I_B\}$ acting only on Alice's Hilbert space. Alice and Bob are prepared initially in an arbitrary state ρ, which is in general not a product state. The state of Bob after a measurement is performed on Alice will then be

$$\begin{aligned}
\rho_B' &= \sum_k \text{tr}_A \left\{ M_k^A \rho (M_k^A)^\dagger \right\} \\
&= \text{tr}_A \left\{ \Big(\sum_k (M_k^A)^\dagger M_k^A \Big) \rho \right\} \\
&= \text{tr}_A(\rho) \\
&= \rho_B.
\end{aligned}$$

I was allowed use the cyclic property of the trace here because the operators M_k^A act only on the Hilbert space which we are tracing off. We therefore see from this general argument that as far as Bob can tell, he has no way of knowing a local operation was performed on Alice.

The no-signaling principle is an *extremely* useful tool when thinking about problems in quantum information. I use it all the time as a kind of sanity check: whenever I'm confronted with a question regarding non-local properties of quantum theories, I always ask myself: "would this violate no-signaling?" All one need to remember is that performing *local* operations on Alice has no effect on any *local* properties on Bob.

No-cloning theorem

Related to no-signaling is the notion of no-cloning: *it is impossible to find a unitary U which clones an arbitrary state $|\psi\rangle$*:

$$U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle, \tag{3.46}$$

for all $|\psi\rangle$. Of course, you can cook up a unitary that does this for a single $|\psi\rangle$. But then this won't work on another $|\phi\rangle$. In fact, the theorem can be proven by contradiction. Suppose we were able to find a unitary U such that

$$U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle, \qquad U|\phi\rangle|0\rangle = |\phi\rangle|\phi\rangle.$$

Then, taking the inner product on both sides would give

$$\langle\psi|\phi\rangle = \langle\psi|\phi\rangle^2.$$

But $x = x^2$ only has solutions $x = 0$ or $x = 1$. Thus, either $|\psi\rangle = |\phi\rangle$ or $|\psi\rangle$ is orthogonal to $|\phi\rangle$. Hence, a general cloning device is not possible. We can only close orthogonal states.

The connection between no-signaling and no-cloning can be made as follows. Suppose Alice and Bob share a Bell state $|\Phi^+\rangle$. Bob then makes many copies of his qubit and measures each one. If in all copies he always finds the same state, he would know that Alice had performed a measurement. If he finds 0 and 1 with equal probabilities, he would know she didn't perform a measurement. Thus, we see that if cloning were possible, Alice could use this idea to transmit a message. No-signaling therefore implies no-cloning.

Measurements in the Bell basis

With composite systems one may also perform non-local measurements. In fact, that's one of the coolest things about quantum mechanics. For instance, a popular protocol is to measure on the Bell basis (3.13)-(3.16). That is perform a measurement described by the four projection operators

$$Q_0 = |\Phi^+\rangle\langle\Phi^+|, \qquad Q_1 = |\Psi^+\rangle\langle\Psi^+|, \qquad Q_2 = |\Phi^-\rangle\langle\Phi^-|, \qquad Q_3 = |\Psi^-\rangle\langle\Psi^-|,$$

Non-local measurements of this form may, at first, seem quite strange. However, they are actually quite simple the perform. All we need to do is use Eq. (3.19) to construct Bell states from the computational basis.

The idea goes as follows. Suppose we have some composite state ρ_{AB} we wish to measure. We then first apply a sequence of two unitaries to the state ρ_{AB}: a CNOT gate U_{CNOT} [Eq. (3.17)] followed by a Hadamard gate H_A [Eq. (3.18)] acting only on A. Afterwards, we measure A and B in the usual computational basis defined by projectors $P_0 = |00\rangle\langle 00|$, $P_1 = |01\rangle\langle 01|$, $P_2 = |10\rangle\langle 10|$ and $P_3 = |11\rangle\langle 11|$. The probability of the different outcomes will be

$$p_i = \text{tr}\bigg\{P_i(H_A \otimes I_B)U_{\text{CNOT}}\rho_{AB}U_{\text{CNOT}}(H_A \otimes I_B)\bigg\}. \tag{3.47}$$

But because of the relations in Eq. (3.19), it follows that the Bell projectors Q_i are related to the computational basis projectors P_i according to

$$Q_i = U_{\text{CNOT}}(H_A \otimes I_B)P_i(H_A \otimes I_B)U_{\text{CNOT}}. \tag{3.48}$$

Hence, the probabilities p_i become

$$p_i = \text{tr}\Big\{Q_i \rho_{AB} Q_i^\dagger\Big\}, \tag{3.49}$$

which are nothing but the probabilities of finding the qubits in the four Bell states. Thus, to perform measurements in the Bell basis, we simply need to first entangle the two qubits and then measure them in the computational basis.

3.5 Bell's inequality

As we saw in the previous section, if Alice and Bob share an entangled state and Alice performs a measurement on her state, the backaction would cause the state of Bob to be immediately updated. This seem strange because Alice and Bob can be galaxies apart. Even though we showed that this does not violate causality (no-signaling), it is still perplexing that such a thing could exist in nature. In fact, consider the following two states:

$$|\psi\rangle_{AB} = \sqrt{p}|00\rangle + \sqrt{1-p}|11\rangle, \tag{3.50}$$

$$\rho_{AB} = p|00\rangle\langle 00| + (1-p)|11\rangle\langle 11|. \tag{3.51}$$

Both states are correlated, but they are fundamentally different. The first is pure state. There is no uncertainty associated to the global state. And it is also entangled. The second, on the other hand, is just a statistical mixture of two possibilities. It has no entanglement nor any other type of quantum correlations. That being said however, in both cases, the probability that Alice finds her qubit in 0 is p. And, in both cases, if she does find it in zero, then the state of Bob will be updated to $|0\rangle$. So in this sense, it seems these two states behave quite similarly.

The state (3.51) represents our degree of **ignorance** about the configurations of Alice and Bob. We don't know in which configuration they are, $|00\rangle$ or $|11\rangle$. If Alice measures and happens to find out, then we update our information. The state (3.50), on the other hand, contains no ignorance at all. We know exactly which state the two qubits are. According to quantum theory, the randomness associated with the state (3.50) has nothing to do with ignorance. It is *intrinsic*.

This gets us into the idea of **realism** or **ontology**. According to philosophy, an object is real when one can associate properties to it *independent* of observation. A rose is red and that is independent of whether you look or not. As Einstein said, "the moon is still there even if we don't look at it". In the mixture (3.51) realism is preserved. The qubits have the property of being either 0 or 1, it is just us, silly physicists, who don't know in which state they are. A state of the form (3.50), however, is not a realist state. Bob's qubit is not in 0 or 1. It will only collapse to 0 or 1 if we happen to make a measurement.

Realism is intimately related to **locality**. If properties can only be established when a system is measured, then measurements on entangled particles would cause an instantaneous backaction on other particles arbitrarily far away. A non-realist theory, like quantum mechanics, must therefore also be intrinsically *non-local*. This led **Einstein Podolsky and Rosen** in 1935[2] to propose that there should exist **past factor** (which, to add some drama, they called *hidden variables*) which determined these properties long ago, perhaps when they interacted. That is, there should be additional variables, perhaps which are very difficult to access in the lab, but which if we knew them, we would know with certainty the configurations of the two qubits. These past factors/hidden variable would therefore recover the status of a realist theory to quantum mechanics. This is the idea behind the **EPR paradox**.

The Bell scenario

For 3 decades the EPR paradox remained almost a philosophical question. There was no way of testing if quantum theory was realist or not. In 1964, however, John Bell proposed a test to do so. Bell's idea was absolutely *brilliant*.

Suppose Alice and Bob are each given one of a pair of (possibly entangled) qubits. They can perform arbitrary measurements on them. That is, they can individually and independently choose the direction

$$\hat{\boldsymbol{n}} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{3.52}$$

in Bloch's sphere, where they wish to measure. Recalling the discussion surrounding Eq. (1.47), what I mean is that they can perform local measurements described by projection operators $P^{\pm}_{\hat{\boldsymbol{n}}} = |\hat{\boldsymbol{n}}_\pm\rangle\langle\hat{\boldsymbol{n}}_\pm|$, where

$$|\boldsymbol{n}_+\rangle = \begin{pmatrix} e^{-i\phi/2}\cos\frac{\theta}{2} \\ e^{i\phi/2}\sin\frac{\theta}{2} \end{pmatrix}, \qquad |\boldsymbol{n}_-\rangle = \begin{pmatrix} -e^{-i\phi/2}\sin\frac{\theta}{2} \\ e^{i\phi/2}\cos\frac{\theta}{2} \end{pmatrix} \tag{3.53}$$

I will also label the outcomes of these measurements as $+1$ and -1 respectively. So in each experiment Alice will either obtain $a = \pm 1$ and Bob will obtain either $b = \pm 1$.

Following the usual literature, I will call x a shorthand for the direction $\hat{\boldsymbol{n}}$ that Alice chooses and y for the choice of Bob. We now repeat the experiments many times. Qubits arrive at Alice and Bob's lab (which are far apart from each other), they choose any direction they want and they measure, obtaining either $+1$ or -1. After collecting a large amount of statistics, we can then reconstruct $P(a, b|x, y)$, which is the probability that Alice gets $a \in \{1, -1\}$ and Bob gets $b \in \{1, -1\}$ *given* that their choice of measurement directions were x and y. According to the rules of quantum mechanics, these probabilities can be constructed as

$$P(a, b|x, y) = \operatorname{tr}\Big\{\rho_{AB}\, P^a_x \otimes P^b_y\Big\}, \tag{3.54}$$

where ρ_{AB} is the global state of the two qubits. However, the idea is precisely *not* to assume that this how these probabilities were computed. Instead, we take a **Black**

---

[2] A. Einstein, B. Podolsky and N. Rosen, *"Can Quantum-Mechanical Description of Physical Reality be Considered Complete?"*, *Phys. Rev.*, **47**, 777780 (1935).

box (device independent) approach: we simply assume that Alice and Bob can make measurements and build from experiment a set of probabilities $P(a, b|x, y)$.

Once we have these probabilities, we can just apply them using the standard tools of statistics. For instance, if we sum over all possible outcomes of Bob, we can construct the marginal distribution of Alice,

$$P(a|x, y) = \sum_b P(a, b|x, y). \tag{3.55}$$

This probability, if physics is to be consistent, should satisfy **no-signaling**:

$$P(a|x, y) = P(a|x). \tag{3.56}$$

That is, the outcomes of Alice cannot be influenced by the *choice* of measurement y that Bob decided to make. That would be super weird (*Quantum Healing is not possible!*).

In general, one will find that the probabilities $P(a, b|x, y)$ are not independent. That is,

$$P(a, b|x, y) \neq P(a|x)P(b|y).$$

The outcomes of Alice and Bob are correlated. But now comes Bell's reasoning.

If these probabilities were constructed by a realist theory, then this independence must stem from a **common source of ignorance**. That is, some **past factor** that established a relation between the two qubits long ago, when they interacted. Let us call this past factor/hidden variable by the generic name λ. What λ is, exactly, is not important. It simply summarizes a set of possible common factors that may have affected the two qubits in the past. If we knew λ, then the outcomes of Alice and Bob would be independent:

$$P(a, b|x, y, \lambda) = P(a|x, \lambda)P(b|y, \lambda). \tag{3.57}$$

The only reason why A and B are correlated is because they share a common ignorance about λ:

$$P(a, b|x, y) = \int \mathrm{d}\lambda\; P(a|x, \lambda)P(b|y, \lambda)q(\lambda), \tag{3.58}$$

where $q(\lambda)$ is the probability distribution of λ. This is the *most general* formulation of a realist theory. Or also, as we mentioned above, a *local* theory since breaking realism must also break locality.

CHSH inequality

The game is now to prove, experimentally, that quantum mechanics can actually *violate* Eq. (3.58). That is, that Nature can produce probability distributions which *cannot* be written in the form (3.58), for any probability distribution $q(\lambda)$.

The easiest way to do that is using an inequality proved by Clauser, Horne, Shimony and Holt in 1969[3] Let us construct expectation values of the possible outcomes. For instance,

$$\langle A_x \rangle = \sum_{ab} a\; P(a, b|x, y).$$

---

[3] J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt, *Phys. Rev. Lett.*, **23** 880 (1969).

This is nothing but the average spin component of Alice's qubit, in the direction x. We can also compute correlation functions, such as

$$\langle A_x B_y \rangle = \sum_{ab} a\, b\, P(a, b|x, y).$$

If we assume that the locality condition (3.58) is true, then we may write

$$\begin{aligned}\langle A_x B_y \rangle &= \sum_{ab} \int \mathrm{d}\lambda\; a\, b\, P(a|x, \lambda) P(b|y, \lambda) q(\lambda)\\ &= \int \mathrm{d}\lambda \langle A_x \rangle_\lambda \langle B_y \rangle_\lambda q(\lambda),\end{aligned}$$

where I defined

$$\langle A_x \rangle_\lambda = \sum_a a\, P(a|x, \lambda),$$

and similarly for $\langle B_y \rangle_\lambda$.

Let us now focus on only two possible directions for Alice and Bob. That is, Alice chooses two directions $\hat{\boldsymbol{n}}$, which we will label as $x = 0, 1$. Similarly, Bob chooses two other directions labeled as $y = 0, 1$. We then construct the following quantity:

$$\boxed{S = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle.} \tag{3.59}$$

If we assume the theory is local, then this may be written as

$$\begin{aligned}S &= \int \mathrm{d}\lambda \Big\{ \langle A_0 \rangle_\lambda \langle B_0 \rangle_\lambda + \langle A_0 \rangle_\lambda \langle B_1 \rangle_\lambda + \langle A_1 \rangle_\lambda \langle B_0 \rangle_\lambda - \langle A_1 \rangle_\lambda \langle B_1 \rangle_\lambda \Big\} q(\lambda)\\ &= \int \mathrm{d}\lambda \Big\{ \langle A_0 \rangle_\lambda \Big[\langle B_0 \rangle_\lambda + \langle B_1 \rangle_\lambda \Big] + \langle A_1 \rangle_\lambda \Big[\langle B_0 \rangle_\lambda - \langle B_1 \rangle_\lambda \Big] \Big\} q(\lambda).\end{aligned}$$

All the expectation values appearing here are bounded to the interval $[-1, 1]$. Hence, the quantity inside {} must lie in the interval $[-2, 2]$. The integral will not change this because $q(\lambda)$ is a probability distribution and convex sums cannot give you something larger than the largest element. Hence, we reach the remarkable conclusion that

$$\boxed{-2 \leq S \leq 2, \qquad \text{(For any local theory).}} \tag{3.60}$$

This is the CHSH inequality.

It is now a matter of pure joy to check that quantum theory can violate this inequality. And if it violates it, then it means quantum theory *cannot* be a realist/local theory.

Violating the CHSH

Suppose the two qubits are prepared in a bell state $|\psi\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. And suppose, for simplicity, that Alice and Bob measure in directions (3.52) with angles θ_x and θ_y and $\phi_x = \phi_y = 0$. Applying Eq. (3.54) one may write, after a bit of simplification,

$$P(a, b|x, y) = \frac{1}{4}\Big\{1 - ab\cos(\theta_x - \theta_y)\Big\}. \tag{3.61}$$

Next choose the directions to be

$$\theta_x \in \{0, \pi/2\}, \qquad \theta_y \in \{-3\pi/4, 3\pi/4\}.$$

One then finds

$$S = 2\sqrt{2} \geq 2. \tag{3.62}$$

Voilá: the inequality is violated. Quantum theory *predicts* non-locality.

If one can now do an experiment which observes a violation of Bell's inequality, then it would mean **Nature** itself, is indeed non-local. And that is exactly what was done. Experiments showing violations of Bell's inequality were first reported by Aspect, Grangier and Roger in 1982. These experiments had certain *loopholes*, however. For instance, Alice and Bob's station were not sufficiently far from each other to forbid classical communication (at the speed of light). Since then, tons of experiments have been made attempting to close these loopholes. An experiment finally closing all of them was reported in 2015.[4] If you are interested in going deeper into the physics of Bell non-locality, I recommend the review by Brunner *et. al.*[5]

---

[4] B. Hansen, *et. al.*, "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres", *Nature*, **526**, 682686 (2015).

[5] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani and S. Wehner, "Bell nonlocality". *Rev. Mod. Phys*, **86(2)**, 419478 (2014).

Chapter 4

Quantifying correlations between quantum systems

4.1 Singular value and Schmidt decompositions

In this section we will have a first look at a long discussion on how to characterize entanglement. Here we will focus on the case where a bipartite system AB is in a pure state $|\psi\rangle_{AB}$. In this case the theory is simple and elegant. The case where AB are already in a mixed state ρ_{AB} is more complicated and will be discussed next chapter.

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

$$|\psi\rangle = \sum_{a,b} \psi_{ab} |a, b\rangle, \tag{4.1}$$

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

$$\rho_A = \operatorname{tr}_B |\psi\rangle\langle\psi| = \sum_{a,a'} \bigg[\sum_b \psi^*_{ab} \psi_{a'b} \bigg] |a\rangle\langle a'|, \tag{4.2}$$

$$\rho_B = \operatorname{tr}_A |\psi\rangle\langle\psi| = \sum_{b,b'} \bigg[\sum_a \psi^*_{ab} \psi_{ab'} \bigg] |b\rangle\langle b'|. \tag{4.3}$$

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 (4.1) that $|\psi\rangle$ will also factor as a product. Our goal will be to quantify the degree of entanglement between these states.

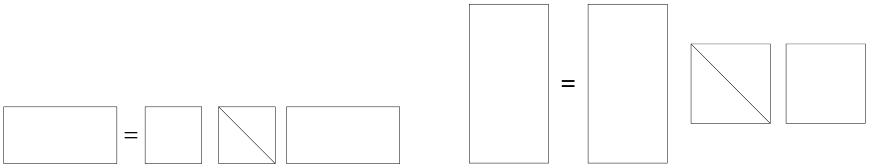

Figure 4.1: The size of the matrices appearing in Eq. (4.4). Left: A is short and fat ($M \leq N$). Right: A is thin and tall ($M \geq N$).

Singular Value Decomposition

To start I will introduce an important tool in linear algebra, 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

$$\boxed{A = USV^\dagger,} \tag{4.4}$$

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, $VV^\dagger = 1$.
- 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. 4.1. For future reference, I will also write down Eq. (4.4) in terms of the components of A:

$$A_{ij} = \sum_{\alpha=1}^{r} U_{i\alpha} \sigma_\alpha V^*_{j\alpha} \tag{4.5}$$

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 the 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. (4.4) and using the fact that $U^\dagger U = 1$ we see that

$$A^\dagger A = VS^2V^\dagger \tag{4.6}$$

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

$$AA^\dagger = US^2U^\dagger \tag{4.7}$$

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 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. (4.5), 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. (4.5) can be written as

$$A = \sum_{\alpha=1}^{r} \sigma_\alpha \boldsymbol{u}_\alpha \boldsymbol{v}_\alpha^\dagger. \tag{4.8}$$

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:

$$A' = \sum_{\alpha=1}^{r'} \sigma_\alpha \boldsymbol{u}_\alpha \boldsymbol{v}_\alpha^\dagger. \tag{4.9}$$

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

$$|\psi\rangle = \sum_{a,b} \psi_{ab} |a, b\rangle. \tag{4.10}$$

With a moment of though we see that ψ_{ab} can also be interpreted as a matrix of coefficients. In fact, this matrix will be rectangular in general, with dimensions $d_A \times d_B$. Interpreting the entries of the vector (4.10) as a matrix is the key idea behind the *Schmidt decomposition*.

I know this is confusing at first. But this is just a reordering of elements. For instance, the state

$$|\psi\rangle = \sqrt{p}|01\rangle + \sqrt{1-p}|10\rangle,$$

can also be represented as a matrix with coefficients $\psi_{01} = \sqrt{p}$ and $\psi_{10} = \sqrt{1-p}$. The same idea also applies if A and B have different dimensions. For instance, suppose A is a qubit but B is a qutrit (3 levels) and they are both prepared in a state of the form

$$|\psi\rangle = \alpha|00\rangle + \beta|02\rangle + \gamma|11\rangle.$$

The corresponding matrix ψ_{ab} will then be

$$\psi = \begin{pmatrix} \alpha & 0 & \beta \\ 0 & \gamma & 0 \end{pmatrix},$$

which has dimensions 2 × 3.

The idea will now be to compute the SVD of the matrix ψ_{ab}. Why this is useful, of course, will become clear in a second. But it is interesting to note that this works in general, even if ψ_{ab} is rectangular. Using Eq. (4.5) we see that this decomposition will have the form

$$\psi_{ab} = \sum_\alpha \sigma_\alpha U_{a\alpha} V^*_{b\alpha}. \tag{4.11}$$

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

$$\sum_{\alpha=1}^{r} \sigma_\alpha^2 = 1. \tag{4.12}$$

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

Inserting Eq. (4.11) back into (4.10) now gives

$$|\psi\rangle = \sum_{a,b,\alpha} \sigma_\alpha U_{a\alpha} V^*_{b\alpha} |a,b\rangle = \sum_\alpha \sigma_\alpha \Big[\sum_a U_{a\alpha}|a\rangle \Big] \otimes \Big[\sum_b V^*_{b\alpha}|b\rangle \Big]. \tag{4.13}$$

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

$$|\alpha_A\rangle = \sum_a U_{a\alpha}|a\rangle, \tag{4.14}$$

$$|\alpha_B\rangle = \sum_b V^*_{b\alpha}|b\rangle. \tag{4.15}$$

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

$$\boxed{|\psi\rangle = \sum_{\alpha} \sigma_\alpha |\alpha_A\rangle \otimes |\alpha_B\rangle.} \tag{4.16}$$

This is way better than (4.10) 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. (3.21) when we first introduced the connection between mixed states and entanglement. The step in going from a general entangled state (4.10) to a state of the form (4.16) 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 (4.10) 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 (4.16), 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 (4.16). Since the states $|\alpha_A\rangle$ and $|\alpha_B\rangle$ are orthonormal, it is straightforward to find that

$$\rho_A = \sum_{\alpha} \sigma_\alpha^2 |\alpha_A\rangle\langle\alpha_A|, \tag{4.17}$$

$$\rho_B = \sum_{\alpha} \sigma_\alpha^2 |\alpha_B\rangle\langle\alpha_B|. \tag{4.18}$$

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

$$\boxed{\mathrm{tr}(\rho_A^2) = \mathrm{tr}(\rho_B^2) = \sum_{\alpha} \sigma_\alpha^4 = \sum_{\alpha} \lambda_\alpha^2.} \tag{4.19}$$

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.

Maximally entangled states

To summarize, I want to emphasize that *all* entanglement properties of a pure state are contained in 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. (4.19). In particular, we now finally have the tools to define what is a *maximally entangled state*: it is a state in which all singular values are equal. Due to the normalization (4.12), this then implies

$$\sigma_\alpha = \frac{1}{\sqrt{r}} \qquad \text{(Maximally entangled states)} \tag{4.20}$$

As an example, consider a state of the form

$$|\psi\rangle = \sqrt{p}|01\rangle + \sqrt{1-p}|10\rangle, \qquad p \in [0,1]. \tag{4.21}$$

We already know that if $p = 0, 1$ the state is a product and if $p = 1/2$ it is a Bell state (3.15). In this case the matrix ψ_{ab} will be

$$\psi = \begin{pmatrix} 0 & \sqrt{p} \\ \sqrt{1-p} & 0 \end{pmatrix}.$$

Recall from Eq. (4.6) that the singular values of ψ are related to the square root of the eigenvalues of $\psi^\dagger\psi$. Well, in our case this is pretty easy because

$$\psi^\dagger\psi = \begin{pmatrix} 1-p & 0 \\ 0 & p \end{pmatrix}.$$

Whence, the singular values are simply

$$\sigma_1 = \sqrt{p}, \qquad \sigma_2 = \sqrt{1-p}. \tag{4.22}$$

When $p = 0, 1$ we have one of the singular values equal to 1, which is the case of a product state. Conversely, when $p = 1/2$ both singular values are equal and we have a maximally entangled state. Thus, the Bell state (3.15) is the maximally entangled state.

A worked example

Let us work out an example in detail. Consider two qubits prepared in the state

$$|\psi\rangle = \frac{c}{\sqrt{2}}(|00\rangle + |11\rangle) + \frac{d}{\sqrt{2}}(|01\rangle + |10\rangle), \tag{4.23}$$

where c and d are constants satisfying $|c|^2 + |d|^2 = 1$. For simplicity I will assume they are real, so that they may be parametrized as $c = \cos(\theta/2)$ and $d = \sin(\theta/2)$. If $(c, d) = (1, 0)$ or $(0, 1)$ we recover the Bell states $|\Phi^+\rangle$ and $|\Psi^+\rangle$ in Eqs. (3.13) and (3.14), which we know are entangled. Conversely, if $c = d = 1/\sqrt{2}$ we actually obtain the product state $|\psi\rangle = |+\rangle|+\rangle$.

The reduced density matrices of A and B are

$$\rho_A = \rho_B = \frac{1}{2}\begin{pmatrix} 1 & 2cd \\ 2cd & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & \sin\theta \\ \sin\theta & 1 \end{pmatrix}. \tag{4.24}$$

The Purity of the global state $\rho_{AB} = |\psi\rangle\langle\psi|$ is of course equal to 1, $\operatorname{tr}\rho_{AB}^2 = 1$, since the state is pure. But in general, the reduced density matrices ρ_A and ρ_B will be mixed. As we just learned from Schmidt's theorem, both will have the same purity,

$$\operatorname{tr}\rho_A^2 = \operatorname{tr}\rho_B^2 = \frac{1+\sin^2\theta}{2}. \tag{4.25}$$

It is true that in this case $\rho_A \equiv \rho_B$, but even in cases when $\rho_A \neq \rho_B$, both will still have the same purity [Eq. (4.19)]. We see that the state is maximally mixed (purity 1/2) when $\theta = 0, \pi$ and that the purity is 1 when $\theta = \pi/2$.

The matrix ψ_{ab} corresponding to the state (4.23) is

$$\psi = \frac{1}{\sqrt{2}}\begin{pmatrix} c & d \\ d & c \end{pmatrix}. \tag{4.26}$$

If you try to plug this on Mathematica to get the Singular Value Decomposition, you may get a bit frustrated. SVD routines are really optimized for numerics and can give weird results for symbolics. In this case, however, the SVD (4.4) is pretty simple. I will leave for you to verify that both U and V in this case are nothing but the Hadamard gate.

$$U = V = H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{4.27}$$

That is,

$$\psi = HSH^\dagger, \tag{4.28}$$

where $S = \operatorname{diag}(\sigma_+, \sigma_-)$, with the singular values

$$\sigma_+ = \frac{c+d}{\sqrt{2}}, \qquad \sigma_- = \frac{c-d}{\sqrt{2}}. \tag{4.29}$$

Applying the recipe (4.19) should give you back Eq. (4.25).

Let us now find the Schmidt basis in Eqs. (4.14) and (4.15). Since $U = V = H$, in this case they are nothing but the $|\pm\rangle$ states. In fact, we may simply verify that the original state (4.23) can actually be written as

$$|\psi\rangle = \frac{(c+d)}{\sqrt{2}}|++\rangle + \frac{(c-d)}{\sqrt{2}}|--\rangle. \tag{4.30}$$

This is the Schmidt form of the state. I personally think it is quite nice. Note how, when written like this, it becomes much more evident when the state will be entangled or not, compared to Eq. (4.23).

Entanglement in mixed states

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 (3.41). 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. We will get back to this topic soon.

Multipartite entanglement and Tensor Networks

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

$$|\psi\rangle = \sum_{s_1,\ldots,s_N} \psi_{s_1\ldots s_N} |s_1, \ldots, s_N\rangle. \tag{4.31}$$

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 $\psi_{s_1,\ldots,s_N}$. And its a mess.

Understanding this mess is a big topic of current research. To complicate things further, one has to notice that if the local Hilbert spaces have dimension d (e.g. $d = 2$ for qubits), then there will be d^N different entries in the state (4.31). 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 silly qubits, we are already in deep trouble. This is not a computational limitation that will be solved with the next generation of processors. It is a *fundamental* limitation.

However, this does not mean that all corners of Hilbert space are equally explored. It may very well be that, in practice, all the action occurs essentially in a small part of it. One is therefore set with the question of understanding which are the relevant corners of Hilbert space.

A particularly fruitful approach to deal with these questions, which has gained enormous popularity in recent years, is to interpret $\psi_{s_1,\ldots,s_N}$ as a rank-N **tensor**. This is the idea of **Tensor Networks**. The neat thing about tensors, is that one can employ low-rank approximations. This is what was done, for instance, when we discussed the SVD as an approximation tool in Eq. (4.9): matrix A_{ij} has N^2 entries, but a matrix of the form $A_{ij} = u_i v_j$ will only have $2N$. Approximating a general A as a sum of $u_i v_j$ allows us to greatly reduce the number of entries. Another reason why Tensor Networks are popular is that they allow one to draw diagrams that tell you who is entangled with whom. 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.

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 **state purification is not unique**. The system B can have any size so 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

$$|\psi\rangle = \sum_a \sqrt{p_a}\; |a\rangle \otimes |a\rangle \tag{4.32}$$

Tracing over B we get

$$\mathrm{tr}_B\, |\psi\rangle\langle\psi| = \rho \tag{4.33}$$

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.

4.2 Mutual information

In the previous section we saw how the Schmidt decomposition could be used to quantify the amount of correlations in bipartite pure states. Any state that has Schmidt rank larger than 1 is entangled. And the degree of entanglement can be measured by the purity of the reduced states, such as $\rho_A = \mathrm{tr}_B\, |\psi\rangle\langle\psi|$. Quantifying correlations in this case is therefore extremely clean and operationally smooth. Unfortunately the same is not true for more complicate states.

By "complicated" I mean one of two things. We can either try to quantify entanglement in **multipartite** systems (e.g. 3 qubits) or we can study correlations in **mixed states**. In both cases, the question on how to quantify quantum correlations is still very much an open topic of research. In this section I will start by discussing the topics which are more or less well established and then afterwards show you some of the most prominent current research directions.

I want to start with a very general kind of quantifier, which is by far the most useful. Let us recall our discussion on von Neumann's entropy in Sec. 2.8. Given a quantum state ρ, the von Neumann entropy is defined as

$$S(\rho) = -\,\mathrm{tr}(\rho \ln \rho). \tag{4.34}$$

The entropy defines the lack of information we have about a quantum state. It is zero when the state is pure (because in this case we have all information – we know exactly what the state is). Moreover, it is $\ln d$ when the state is the identity $\pi = \mathbb{I}/d$ (maximally mixed state). Another quantity, intimately related to $S(\rho)$, is the relative entropy (Kullback-Leibler divergence). Given two density matrices ρ and σ, it is defined as

$$S(\rho||\sigma) = \mathrm{tr}\Big\{\rho \ln \rho - \rho \ln \sigma\Big\}. \tag{4.35}$$

The relative entropy represents a kind of distance between ρ and σ. It satisfies $S(\rho||\sigma) \geq 0$, with equality if and only if $\rho = \sigma$. To make the link between the von Neumann entropy and the relative entropy, we compute the relative entropy between a state ρ and the maximally mixed state $\pi = \mathbb{I}/d$. It reads $S(\rho||\pi) = -S(\rho) + \ln d$. This motivates us to define the information we have about a quantum state as

$$\mathcal{I}(\rho) = S(\rho||\pi) = \log d - S(\rho). \tag{4.36}$$

The information is therefore the distance between ρ and the maximally mixed state. Makes sense.

The logic behind Eq. (4.36) is extremely powerful and is the basis for most quantifiers in quantum information theory. In (4.36) we wish to quantify information, so we consider the distance between our state and a reference state, which is the state for which we have no information.

We use this idea to introduce the concept of **Mutual Information**. Consider a bipartite system AB prepared in an arbitrary state ρ_{AB}. This state will in general not be a product, so from the marginals $\rho_A = \mathrm{tr}_B\, \rho_{AB}$ and $\rho_B = \mathrm{tr}_A\, \rho_{AB}$, we cannot reconstruct the original state. That is, $\rho_A \otimes \rho_B \neq \rho_{AB}$. We then ask, what is the information that is contained in ρ_{AB}, but which is *not* present in the marginalized state $\rho_A \otimes \rho_B$. This is a measure of the correlations present in ρ_{AB}. This can be defined as the distance between ρ_{AB} and the marginalized state $\rho_A \otimes \rho_B$:

$$\mathcal{I}_{\rho_{AB}}(A\!:\!B) = S(\rho_{AB}||\rho_A \otimes \rho_B).$$

Due to the properties of the relative entropy, this quantity is non-negative and zero if and only if ρ_{AB} is a product state. For this reason, the mutual information quantifies the amount of correlations in ρ_{AB}. Using Eq. (4.35) we can also write

$$S(\rho_{AB}||\rho_A \otimes \rho_B) = -S(\rho_{AB}) - \mathrm{tr}\Big\{\rho_{AB} \ln \rho_A\Big\} - \mathrm{tr}\Big\{\rho_{AB} \ln \rho_B\Big\}.$$

But now we can compute the partial traces in two steps. For instance,

$$-\,\mathrm{tr}\Big\{\rho_{AB} \ln \rho_A\Big\} = -\,\mathrm{tr}_A\Big\{\mathrm{tr}_B\Big[\rho_{AB}\Big] \ln \rho_A\Big\} = -\,\mathrm{tr}_A\Big\{\rho_A \ln \rho_A\Big\}.$$

Hence, we find that the mutual information may also be written as

$$\boxed{\mathcal{I}_{\rho_{AB}}(A\!:\!B) = S(\rho_{AB}||\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}).} \tag{4.37}$$

The mutual information quantifies the total amount of correlations present in a state, irrespective of whether these correlations are classical or quantum. It is, by far, the most important quantifier of correlations in quantum information theory. It is also extremely easy to use. A harder question to address is *"how much of $\mathcal{I}_{\rho_{AB}}(A:B)$ is quantum and how much is classical."* That is a big question and still an open topic of research. We will talk more about it below. If the state ρ_{AB} is pure than $S(\rho_{AB}) = 0$. Moreover, as we saw in Eqs. (4.17) and (4.18) of the previous section, the eigenvalues of ρ_A and ρ_B in this case will be the same. Hence $S(\rho_A) = S(\rho_B)$. We therefore conclude that, for pure states,

$$\mathcal{I}_{\rho_{AB}}(A:B) = 2S(\rho_A) = 2S(\rho_B), \qquad \text{(for pure states)}. \tag{4.38}$$

When the state is pure, all correlations are quantum and correspond to entanglement. Hence, we see that in this case the mutual information becomes twice the entanglement entropy.

As a byproduct of Eq. (4.37), since $\mathcal{I}_{\rho_{AB}}(A:B) \geq 0$, we also learn that

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B). \tag{4.39}$$

This is called the **subadditivity** of the von Neumann entropy: the entropy of the whole is always lesser or equal than the entropy of the parts.

Next consider now the information (4.36) of a quantum state ρ_{AB} (not the mutual, just the standard information). It reads $\mathcal{I}(\rho_{AB}) = \log d_A d_B - S(\rho_{AB})$. Using Eq. (4.37) we can write this as

$$\mathcal{I}(\rho_{AB}) = \mathcal{I}(\rho_A) + \mathcal{I}(\rho_B) + \mathcal{I}_{\rho_{AB}}(A:B). \tag{4.40}$$

This makes the physical meaning of the mutual information particularly clear: The total information $\mathcal{I}(\rho_{AB})$ contained in the state ρ_{AB} is split into the local information, contained in the marginals ρ_A and ρ_B, plus the *mutual* information that is shared between A and B.

As we just saw, the mutual information is a distance between a state ρ_{AB} and the marginalized state $\rho_A \otimes \rho_B$. We can actually make this a bit more general and define the mutual information as the *minimum* distance between ρ_{AB} and *all* product states,

$$\mathcal{I}_{\rho_{AB}}(A:B) = \min_{\sigma_A,\sigma_B} S(\rho_{AB}||\sigma_A \otimes \sigma_B), \tag{4.41}$$

where σ_A and σ_B are arbitrary states of A and B. It is easy to convince ourselves that the closest product state to ρ_{AB} is actually $\rho_A \otimes \rho_B$. To do this, we use the definition of the relative entropy in Eq. (4.35) together with the fact that ρ_A and ρ_B are the marginals of ρ_{AB}, to write

$$S(\rho_{AB}||\sigma_A \otimes \sigma_B) = S(\rho_{AB}||\rho_A \otimes \rho_B) + S(\rho_A \otimes \rho_B||\sigma_A \otimes \sigma_B) \geq S(\rho_{AB}||\rho_A \otimes \rho_B),$$

since the relative entropy is non-negative. Hence, the minimization clearly occurs for $\sigma_{A(B)} = \rho_{A(B)}$. We thus see that out of the set of all product states, that which is closes to ρ_{AB} ("closest" in the sense of the relative entropy) is exactly $\rho_A \otimes \rho_B$.

4.3 Other quantifiers based on relative entropies

A very nice idea

The reason why it is interesting to write the mutual information as the minimization in Eq. (4.41) is because it introduces an *extremely* powerful idea:[1] suppose we want to quantify some property $\mathcal{P}$ of a quantum state ρ. Any property. We then first define the set $\mathcal{S}$ of all states which do not have that property. The distance between ρ and the closest state *without* that property can then be used to quantify that property:

$$\mathcal{P}(\rho) = \min_{\sigma\in\mathcal{S}} S(\rho||\sigma). \tag{4.42}$$

This is one of the basic ideas behind **quantum resource theories**. We say a property $\mathcal{P}$ is a resource. And we establish what are the set of states which do not have that resource (the **free states**). The amount of resource that a state ρ has can then be associated with the minimum distance between ρ and the set of resourceless states. If this distance is large, then ρ will have a substantial amount of this resource. There is also another essential piece to resource theories, which are the set of operations one can perform and which do not increase a resource. We will talk more about this when we discuss quantum operations in general, next chapter.

The relative entropy of coherence

To give you an example, consider a specific basic $|i\rangle$ of a quantum system. And let us suppose that the property we wish to quantify (our resource) is how much coherence a state has with respect to this basis. We then first define the set $\mathcal{I}$ of *incoherent states* (with respect to the basis $|i\rangle$). This will be the set of all states which are diagonal in this basis,

$$\delta = \sum_i \delta_i |i\rangle\langle i| \in \mathcal{I}. \tag{4.43}$$

Given a state ρ, we can now quantify the amount of coherence present in it by

$$C(\rho) = \min_{\delta\in\mathcal{I}} S(\rho||\delta).$$

This is called the **relative entropy of coherence**.[2] In this case the minimization can again be done in closed form. Given a general $\rho = \sum\limits_{ij} \rho_{ij}|i\rangle\langle j|$, define $\rho_{\text{diag}} = \sum\limits_i \rho_{ii}|i\rangle\langle i|$ as the state containing only the diagonal entries of ρ in the basis $|i\rangle$. One may then verify that for any $\delta\in\mathcal{I}$, $\text{tr}(\rho\log\delta) = \text{tr}(\rho_{\text{diag}}\log\delta)$. This allows us to write

$$S(\rho||\delta) = S(\rho_{\text{diag}}) - S(\rho) + S(\rho_{\text{diag}}||\delta) \geq S(\rho_{\text{diag}}) - S(\rho).$$

---

[1] See K. Modi, *et. al.*, “Unified view of quantum and classical correlations”. *Phys. Rev. Lett.*, 104, 080501 (2010).

[2] T. Baumgratz, M. Cramer, and M. B. Plenio, “Quantifying coherence”, *Phys. Rev. Lett.*, **113**, 140401. (2014).

Hence, the minimum occurs precisely for $\delta = \rho_{\text{diag}}$, so that the relative entropy of coherence may be written as

$$\boxed{C(\rho) = \min_{\delta \in \mathcal{I}} S(\rho \| \delta) = S(\rho_{\text{diag}}) - S(\rho).} \tag{4.44}$$

This quantity serves as a faithful measure of the amount of coherences in a basis $|i\rangle$, present in a quantum state ρ.

Separable states and LOCC

Ok. Let's go back to quantum correlations. A bipartite state ρ_{AB} is said to be **separable** if it can be written in the form

$$\rho_{AB} = \sum_i p_i \, \rho_A^i \otimes \rho_B^i, \tag{4.45}$$

where ρ_A^i and ρ_B^i are valid density matrices for systems A and B. This is a typical state mixing classical and quantum stuff. The states ρ_A^i and ρ_B^i can be anything we want. But we are mixing them in a classical way. With some classical probability p_1 system A is prepared in ρ_A^1 and B in ρ_B^1, with a probability p_2 they are prepared in ρ_A^2 and ρ_B^2, and so on. Thus, even though quantumness may be hidden inside the ρ_A^i and the ρ_B^i, the correlations between A and B are purely classical, being related simply to a lack of information about which preparation took place.

It can be shown that separable states can be produced solely with *Local Operations and Classical Communications (LOCC)*. That is, they can be produced by doing stuff locally in A and B, as well as sending WhatsApp messages. For this reason, *separable states are said not to be entangled*. That is, we define entanglement for mixed states as states which are not separable. Following the spirit of Eq. (4.42), the amount of entanglement in a quantum state ρ_{AB} can be quantified by the **relative entropy of entanglement**:

$$\mathcal{E}(\rho_{AB}) = \min_{\sigma_{AB} \in \mathcal{S}} S(\rho_{AB} \| \sigma_{AB}), \tag{4.46}$$

where $\mathcal{S}$ is the set of separable states. Unfortunately, no closed form exists for $\mathcal{E}(\rho_{AB})$. And, what is worse, this quantity is notoriously difficult to compute. In fact, this is considered one of the major open problems in quantum information theory.

This is frustrating since entanglement is a key resource in quantum information processing applications. An alternative way to quantify entanglement using what is called the **Entanglement of Formation (EoF)**. This at least has a closed (albeit ugly) formula for two qubits. The idea is as follows. We know how to quantify entanglement for pure states. In this case the entanglement entropy is simply the von Neumann entropy of the reduced states,

$$\mathcal{E}(|\psi\rangle\langle\psi|) = S(\rho_A), \qquad \rho_A = \text{tr}_B \, |\psi\rangle\langle\psi|. \tag{4.47}$$

Given a mixed state ρ_{AB}, we therefore look into all possible decompositions of it in terms of an *ensemble of pure states*. That is, decompositions of the form

$$\rho_{AB} = \sum_i q_i |\psi_i\rangle\langle\psi_i|. \tag{4.48}$$

where q_i are probabilities and $|\psi_i\rangle$ are arbitrary states. The entanglement of formation is then defined as

$$\mathcal{E}_F(\rho_{AB}) = \min \sum_i q_i \, \mathcal{E}(|\psi_i\rangle\langle\psi_i|), \tag{4.49}$$

where the minimization is done over all possible decompositions into ensembles of pure states.

For the case of two qubits, one can find a closed formula for $\mathcal{E}(\rho_{AB})$ which reads as follows.[3] I will list it here in case you ever need it in the future. But please, don't let the messiness of the next few lines interrupt your reading.

Define a matrix

$$\tilde{\rho}_{AB} = (\sigma_y \otimes \sigma_y)\rho^*_{AB}(\sigma_y \otimes \sigma_y).$$

where ρ^* is the complex conjugate of ρ. Then define the matrix

$$R = \sqrt{\sqrt{\rho_{AB}}\,\tilde{\rho}_{AB}\,\sqrt{\rho_{AB}}}.$$

(Yeah, this is ugly. I know.) Next let $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4$ denote the four eigenvalues of R in decreasing order. We then define the **Concurrence** as

$$C(\rho_{AB}) = \max\Big\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\Big\}. \tag{4.50}$$

The Concurrence in itself can already be taken as a measure of correlations, being 0 whenever a state is separable and 1 for the case of Bell (maximally entangled) states. The entanglement of formation is then given by

$$\mathcal{E}_F(\rho_{AB}) = H\bigg(\frac{1+\sqrt{1-C^2}}{2}\bigg), \qquad H(x) = -x\log x - (1-x)\log(1-x). \tag{4.51}$$

Although ugly, this formula is still widely used as a quantifier of entanglement. It does not measure the entanglement in the nice distance-based way of Eq. (4.46). Rather, it quantifies the number of Bell pairs required to create a quantum state.

Example: Werner state

As an example, consider the Werner state of two qubits:

$$\rho = p\frac{I}{4} + (1-p)|\Psi^+\rangle\langle\Psi^+|, \tag{4.52}$$

where $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ is the Bell state. The Entanglement of Formation (4.51) for this state is plotted in Fig. 4.2 as a function of p. When $p = 0$ we have just the Bell state so $\mathcal{E}_F = 1$. As p increases, $\mathcal{E}_F$ goes down up to $p = 2/3$, after which $\mathcal{E}_F \equiv 0$. This means the state becomes separable for $p \geq 2/3$.

---

[3] S. Hill, W. K. Wooters, "Entanglement of a Pair of Quantum Bits Scott", *Phys. Rev. Lett.*, **78**, 5022 (1997).

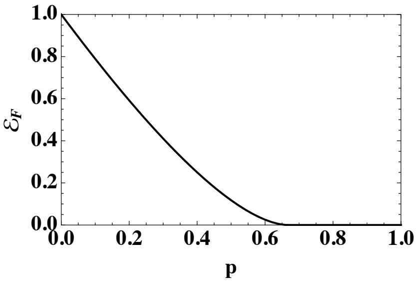

Figure 4.2: Entanglement of Formation $\mathcal{E}_F$ [Eq. (4.51)] as a function of p for the Werner state (4.52).

4.4 Quantum Discord

Classical-classical states

We continue with the discussion on how to quantify the quantum aspects of correlations. This separation of what is quantum and what is classical is not sharp and there are different ways of looking at the problem. As we saw, separable states of the form (4.45) are said not to be entangled because any correlations comes solely from the probabilities p_i that represent a classical uncertainty about the preparation. However, that does not mean that the ρ_A^i and ρ_B^i don't have a lot of quantumness in them.

So let us take another approach. Can we at least agree on a state which, although correlated, can be considered to be *fully classical*, in absolutely all senses imaginable? Yes, it is a state of the form

$$\rho_{AB} = \sum_{ij} p_{ij} |ij\rangle\langle ij|, \tag{4.53}$$

where p_{ij} are probabilities and $|i\rangle_A$ and $|j\rangle_B$ represent arbitrary bases of A and B. Why is this state classical? Because it satisfies absolutely all properties expected from classical probability theory. Or, put it differently, for a state like this all we need is the joint probability distribution p_{ij}. From it we can deduce any other property.

Let me try to explain what I mean. In classical probability theory if you are given a joint probability distribution p_{ij} of A and B, one can compute the marginal distributions by simply summing over the undesired index:

$$p_i^A = \sum_j p_{ij}, \qquad p_j^B = \sum_i p_{ij}. \tag{4.54}$$

On the other hand, given a quantum state ρ_{AB}, we can compute the reduced density matrix by taking the partial trace. For instance,

$$\rho_A = \text{tr}_B\, \rho_{AB} = \sum_{ij} p_{ij} |i\rangle\langle i|.$$

But we can also write this as

$$\rho_A = \sum_i \Big[\sum_j p_{ij} \Big] |i\rangle\langle i| = \sum_i p_i^A |i\rangle\langle i|.$$

Thus, we see that taking the partial trace is the same as just computing the marginal distribution p_i^A. We don't even need the quantum state: we can simply operate with a probability distribution.

But there is also a stronger reason as to why the state (4.53) can be considered classical. And it is related to *measurements* and **conditional probability distributions**. In probability theory we define the conditional distribution of A given B as

$$p_{i|j} = \frac{p_{ij}}{p_j^B}. \tag{4.55}$$

This represents the probability of observing outcome i for A *given* that outcome j for B was observed.

Here is where things get interesting, because in quantum mechanics, in order to observe an outcome for B one must perform a measurement on it. And, as we know, measurements usually have a backaction. The state (4.53) precisely because it does *not*. Suppose we measure B in the basis $|j\rangle$. If outcome $|j\rangle$ is observed, the state of AB after the measurement will then be updated to

$$\rho_{AB|j} = \frac{1}{p_j}\Big[I_A \otimes |j\rangle\langle j|\Big]\rho_{AB}\Big[I_A \otimes |j\rangle\langle j|\Big], \tag{4.56}$$

where

$$p_j = \operatorname{tr}\Big[I_A \otimes |j\rangle\langle j|\Big]\rho_{AB}\Big[I_A \otimes |j\rangle\langle j|\Big]. \tag{4.57}$$

Applying this to the state (4.53) yields

$$\Big[I_A \otimes |j\rangle\langle j|\Big]\rho_{AB}\Big[I_A \otimes |j\rangle\langle j|\Big] = \sum_i p_{ij}|ij\rangle\langle ij| \tag{4.58}$$

To compute p_j we take the trace. Not surprisingly, we find that $p_j = p_j^B = \sum_i p_{ij}$, the marginal of B. That is, the probability of finding outcome j when we measure B is simply the marginal distribution p_j^B. The conditional state, given that j was observed also becomes

$$\rho_{AB|j} = \sum_i \frac{p_{ij}}{p_j^B}|ij\rangle\langle ij|.$$

The reduced density matrix of B is simply $\rho_{B|j} = |j\rangle\langle j|$. And, what is the interesting part, the reduced density matrix of A will be

$$\rho_{A|j} = \operatorname{tr}_B \rho_{AB|j} = \sum_i \frac{p_{ij}}{p_j^B}|i\rangle\langle i| = \sum_i p_{i|j}|i\rangle\langle i|. \tag{4.59}$$

We see that this is once again nothing but what one would obtain from the conditional probability distribution (4.55). The basis elements $|i\rangle\langle i|$ aren't really doing anything interesting. They are just sort of hanging in there.

To summarize, the state (4.53) can be considered a genuinely classical state because it satisfies all properties expected from classical probability theory. There exists bases sets $\{Ii\rangle_A\}$ and $\{|j\rangle_B\}$ for which no backaction in the measurements occur, so that the notion of conditional probability can be defined without a problem. The state still has correlations because in general $p_{ij} \neq p_i^A p_j^B$. But these correlations are fully classical. States of the form (4.53) are also called classical-classical states, not because they are really really classical, but because they are classical on both sides A and B. I think it is really nice how the density matrix formalism contains all of classical probability theory. Density matrices can go from purely quantum stuff to purely classical.

Mutual information for classical-classical states

The fact that the state (4.53) is fully classical also means that information-theoretic quantities for it can be written as in classical information theory. For instance,

$$S(\rho_{AB}) = \sum_{ij} p_{ij} \log p_{ij},$$

$$S(\rho_A) = \sum_i p_i^A \log p_i^A,$$

$$S(\rho_B) = \sum_j p_j^B \log p_j^B.$$

The Mutual Information (4.37) can be written in a neat way if we write, for instance,

$$S(\rho_A) = \sum_{ij} p_{ij} \log p_i^A.$$

This formula is true because j only appears in the first p_{ij}, so that if we sum over j we get $\sum_j p_{ij} = p_i^A$. But with this trick we can write all 3 terms in Eq. (4.37) in a single sum, leading to

$$\mathcal{I}(A\!:\!B) = \sum_{ij} p_{ij} \log \frac{p_{ij}}{p_i^A p_j^B} = D(p_{AB} || p_A p_B), \tag{4.60}$$

where $D(p_{AB}||p_A p_B)$ is the classical version of the relative entropy (4.35),

$$D(p||q) = \sum_i p_i \log p_i/q_i. \tag{4.61}$$

This result provides an intuitive feel for the mutual information. It simply quantifies the ratio between the probability distribution p_{ij} and the marginalized distribution $p_i^A p_j^B$.

We can also write it in terms of the conditional distributions $p_{i|j}$ in Eq. (4.55). It then becomes

$$\boxed{\mathcal{I}(A\!:\!B) = \sum_j p_j^B \sum_i p_{i|j} \log \frac{p_{i|j}}{p_i^A} = \sum_j p_j^B\, D(p_{A|j}||p_A).} \tag{4.62}$$

An analogous formula can be written by conditioning over A instead of B. This formula has a very nice interpretation. The mutual information is simply the average of the relative entropy between the distribution of A given j and the distribution of A without knowledge of j. The quantity $p_{i|j}/p_i^A$ represents how much learning something about B affects the outcomes of A.

We can also write (4.62) in a third way. If we split the log in two terms, the one involving p_i^A will simply reduce to

$$-\sum_j p_j^B \sum_i p_{i|j} \log p_i^A = -\sum_i p_i^A \log p_i^A = H(A),$$

This is the **Shannon entropy** H of the distribution p_A. As for the other term, we define the **conditional entropy**

$$H(A|B) = -\sum_j p_j^B \sum_i p_{i|j} \log p_{i|j}. \tag{4.63}$$

The entropy $H(A)$ measures the lack of information we have about A. The conditional entropy $H(A|B)$ measures the lack of information about A given we know what the outcome of B is. In terms of these quantities, the mutual information (4.62) can be written as

$$\boxed{\mathcal{I}(A\!:\!B) = H(A) - H(A|B).} \tag{4.64}$$

Thus, the mutual information is nothing but the difference between how much you don't know about A and how much you don't know about A given that you know B. Of course, again, the results are symmetric in A and B so that it is also true that $\mathcal{I}(A\!:\!B) = H(B) - H(B|A)$.

Quantum-classical states

The results above for the mutual information apply only for the classical-classical state (4.53). Let us now explore what happens when we start to introduce some quantumness in our state. The next best thing would be a classical-quantum state

$$\rho_{AB} = \sum_j \rho_j^A \otimes \Big[p_j^B |j\rangle\langle j|_B\Big] \tag{4.65}$$

where ρ_j^A are arbitrary density operators for A. This state behaves in a simple way with respect to B, but may contain a bunch of quantumness in A. The reduced states are:

$$\rho_A = \sum_j p_j^B \rho_j^A,$$

$$\rho_B = \sum_j p_j^B |j\rangle\langle j|.$$

As a particular case, if $\rho_j^A = \sum_i p_{i|j} |i\rangle\langle i|$, then we recover the classical-classical state (4.53).

To give an example, suppose A and B are qubits with $\rho_0^A = |0\rangle\langle 0|$ and $\rho_1^A = |+\rangle\langle +|$. Then

$$\rho_{AB} = p_0^B |0\rangle\langle 0| \otimes |0\rangle\langle 0| + p_1^B |+\rangle\langle +| \otimes |1\rangle\langle 1|. \tag{4.66}$$

The reduced state of A will then be

$$\rho_A = p_0^B|0\rangle\langle 0| + p_1^B|+\rangle\langle +|.$$

We now try to play with conditional distributions and see what happens. First, suppose we measure B in the basis $|j\rangle$. Following the same steps in Eq. (4.58), we get

$$\Big[I_A \otimes |j\rangle\langle j|\Big]\rho_{AB}\Big[I_A \otimes |j\rangle\langle j|\Big] = \rho_j^A \otimes \Big[p_j^B|j\rangle\langle j|\Big].$$

Thus, the probability of obtaining outcome j is simply p_j^B, as expected. Moreover, the conditional state of A given that j was observed is

$$\rho_{A|j} = \rho_j^A, \tag{4.67}$$

which of course makes sense.

But now we try to measure A instead. Here things become tricky because, in general, it is impossible to find a measurement that does not have a backaction, like in the classical-classical case. This becomes quite clear if we look at the example in Eq. (4.66). If we measure A in the computational basis, there will be a non-trivial backaction in the $|+\rangle\langle +|$ part. And if we measure A in the $|\pm\rangle$ basis, there will be a backaction on the $|0\rangle\langle 0|$ part. Thus, if we do stuff from B's side, we get usual probability theory, but if we do it on A's side, we get quantum behavior. It is important to realize that even in the classical-classical case, one can still have measurement backaction. But in that case there *exists* a choice of measurement for which there is no backaction. In the case of the quantum-classical case such a measurement scheme, in general, does not exist. That is the fundamental difference between both cases.

Quantum discord

Let us now turn to the Mutual Information for the quantum-classical state (4.65). The Mutual Information is of course given by Eq. (4.37). That definition is rock solid and always holds, no matter what type of state. However, we also showed that in the classical-classical state we could write it as (4.64). This way of writing is nice because it relates information to the difference in entropy between the case where you know, and the case where you do not know, the other side. Can we have an equivalent formulation valid for more general states?

The answer is "sort of". First, we need to realize that the type of measurement *matters*. The amount of information we can learn about A given that we measured B, depends on what measurement we did on B. So let us make things a bit more formal by using the concept of generalized measurements studied in Sec. 2.7 and 3.4. Let M_k^B denote a set of measurement operators acting on B, satisfying $\sum_k (M_k^B)^\dagger M_k^B = I$. We then start with an arbitrary state ρ_{AB} and perform the M_k^B measurement on B. The probability of outcome k will be

$$p_k^B = \operatorname{tr}\Big\{(I_A \otimes M_k^B)\rho_{AB}(I_A \otimes M_k^B)^\dagger\Big\}, \tag{4.68}$$

and the state of AB given that outcome k was obtained, will be

$$\rho_{AB|k} = \frac{(I_A \otimes M_k^B)\rho_{AB}(I_A \otimes M_k^B)^\dagger}{p_k}. \tag{4.69}$$

From this we compute the reduced state of A, $\rho_{A|k}$. We then define the **conditional entropy**

$$S_M(A|B) = \sum_k p_k^B S(\rho_{A|k}). \tag{4.70}$$

I write this as S_M to emphasize that it depends on the choice of measurement $\{M_k^B\}$. This quantity represents the total lack of knowledge on the state of A given that that we know the outcomes of the M_k^B measurement.

Finally, from this we can now define

$$\boxed{\mathcal{J}_M(A|B) = S(\rho_A) - S_M(A|B).} \tag{4.71}$$

This quantity represents how much the information about A was updated given the knowledge we learned from measuring B. It is the generalization of Eq. (4.64) for arbitrary quantum states.

The key point however, which we will now explore, is that for general quantum states $J_M(A|B)$ is in general *not* the same as the mutual information $\mathcal{I}(A:B)$. Their difference is called **Quantum Discord**:

$$\boxed{Q_M(A|B) = \mathcal{I}(A:B) - J_M(A|B).} \tag{4.72}$$

The discord is always non-negative (as we will show in a second). It represents the mismatch between the total amount of shared information between A and B and the information gain related to the specific choice of measurement M_k^B. There is a "Discord" because in quantum theory the type of measurement *matters* and may not give you the full information.

In fact, we can even make that more rigorous: we can define a *measurement independent discord* by minimizing the distance (4.72) over all possible measurements:

$$\boxed{Q(A|B) = \min_{M_k^B} Q_M(A|B).} \tag{4.73}$$

The remarkable feature of quantum mechanics is that there exists states for which this quantity is non-zero, meaning no measurement is capable of recovering the results expected from classical probability theory. Computing this minimization is not an easy task, which is a bummer. And also, sometimes the measurement which extracts the largest information is actually quite crazy and would be impossible to perform in a lab. But nevertheless, you have to agree with me that conceptually the idea is quite beautiful.

Discord in quantum-classical states

Let us now go back to the quantum-classical state (4.65) and explore these new ideas. We begin by proving something called the **joint entropy theorem**. Consider the spectral decomposition of each ρ_j^A in the form

$$\rho_j^A = \sum_k p_{k|j}^A |k_j\rangle\langle k_j|.$$

Remember that the ρ_j^A are just arbitrary density matrices of A. Thus, for the same j, the $|k_j\rangle$ form an orthonormal basis, $\langle k_j|q_j\rangle = \delta_{k,q}$. But for different j there is no particular relation between them, $\langle k_j|q_{j'}\rangle \neq \delta_{k,q}$. The quantum-classical state (4.65) can then be written as

$$\rho_{AB} = \sum_{k,j} p_{k|j}^A p_j^B |k_j, j\rangle\langle k_j, j|. \tag{4.74}$$

The states $|k_j, j\rangle$ form an orthonormal basis since

$$\langle k_j, j|q_{j'}, j'\rangle = \delta_{j,j}\langle k_j|q_j\rangle = \delta_{j,j'}\delta_{k,q}.$$

Thus, we conclude that the eigenvalues of ρ_{AB} are simply $p_{k|j}^A p_j^B$.

The von Neumann entropy of ρ_{AB} will then be

$$S(\rho_{AB}) = -\operatorname{tr}\rho_{AB}\log\rho_{AB} = -\sum_{k,j} p_{k|j}^A p_j^B \log p_{k|j}^A p_j^B.$$

Splitting the log in two terms allows us to identify known structures:

$$-\sum_{k,j} p_{k|j}^A p_j^B \log p_j^B = -\sum_j p_j^B \log p_j^B = H(p_B),$$

$$-\sum_{k,j} p_{k|j}^A p_j^B \log p_{k|j}^A = \sum_j p_j^B S(\rho_j^A).$$

We thus arrive at the **joint entropy theorem**,

$$\boxed{S\Big(\sum_j \rho_j^A \otimes [p_j^B |j\rangle\langle j|_B]\Big) = H(p_B) + \sum_j p_j^B S(\rho_j^A).} \tag{4.75}$$

The connection with quantum discord can now be made clear. The mutual information (4.37) for the quantum-classical state becomes

$$\begin{aligned}\mathcal{I}(A\!:\!B) &= S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \\ &= S(\rho_A) + H(p_B) - H(p_B) - \sum_j p_j^B S(\rho_j^B) \\ &= S(\rho_A) - -\sum_j p_j^B S(\rho_j^B)\end{aligned}$$

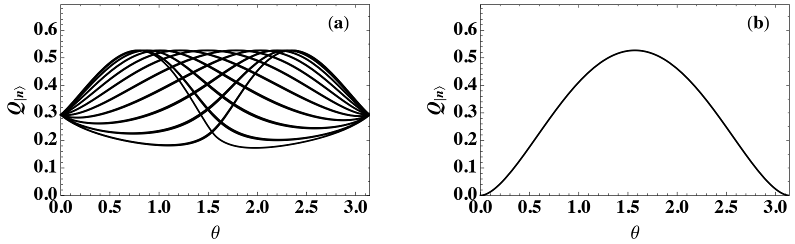

Figure 4.3: Basis dependent quantum discord (4.72) for the quantum-classical state (4.66) and a projective measurement performed in the basis (4.76). (a) When the measurement is performed in A. (b) Measurement performed in B.

This is nothing but $J_{|j\rangle}(A|B)$ in Eq. (4.71). Thus, we see that for a quantum-classical state, if we measure B in the basis $|j\rangle$ there will be no discord at all. Of course, if we measure B in some crazy basis, there will be a discord. But at least there exists one basis for which the minimization in Eq. (4.73) gives exactly zero.

The discord, however, is not symmetric in A and B. That is the whole point of having a quantum-classical state. So even though there exists zero-discord measurements on B, in general this is not the case for measurements performed on A. That is, even though we may have $\mathcal{Q}(A|B) = 0$, we may very well have $\mathcal{Q}(B|A) \neq 0$.

To illustrate this, let us consider the quantum-classical state in Eq. (4.66) and let us suppose that we perform projective measurements on each qubit in the basis

$$|\boldsymbol{n}_+\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle, \qquad |\boldsymbol{n}_-\rangle = -e^{-i\phi}\sin(\theta/2)|0\rangle + \cos(\theta/2)|1\rangle. \quad (4.76)$$

In Fig. 4.3 we show the basis dependent discord (4.72) as a function of θ for the case where the measurement is performed in A and in B. The curves suppose $p = 0.3$. In the case of measurements in B the discord turns out not to depend on ϕ. The curve for when the measurement is performed in A [Fig. 4.3(a)] represent multiple values of ϕ.

As can be seen, when the measurement is performed in B, which is the classical side of the quantum-classical state, there exist values of θ for which the discord is zero. This is precise $\theta = 0$, which means measuring in the computational basis. Conversely, when the measurement is performed in A, no matter what measurement direction (θ, ϕ) we choose, we always get a non-zero discord.

Chapter 5

Open quantum systems

5.1 Overview of quantum operations

The evolution of closed systems is described by unitary operations

$$\rho' = U\rho U^\dagger, \qquad U^\dagger U = 1. \tag{5.1}$$

As we already saw in previous chapters, this is not the most general type of transformation taking density matrices to density matrices. The most general map is called a **quantum operation** or **quantum channel**. It has the form

$$\rho' = \mathcal{E}(\rho) = \sum_k M_k \rho M_k^\dagger, \tag{5.2}$$

where $\{M_k\}$ is a set of operators satisfying the condition

$$\sum_k M_k^\dagger M_k = 1, \tag{5.3}$$

known as **Kraus operators**. If the set $\{M_k\}$ has only one operator, we are back to Eq. (5.1). The map described by Eq. (5.2) is called **Completely Positive and Trace Preserving (CPTP)**. The trace preserving condition is ensured by Eq. (5.3). The term "completely positive" means that it takes positive operators to positive operators.

There are four names in this chapter: Kraus, Stinespring, Lindblad and Choi. Here is what each name is associated with:

- **Kraus:** the representation of a channel as in Eq. (5.2).
- **Stinespring:** a channel of the form (5.2) can always be viewed as the interaction of a system with an environment (hence justifying the name *open quantum systems*).
- **Lindblad:** a way to write the map (5.2) as a differential equation $d\rho/dt = \ldots$. This is a bit like how von Neumann's equation is derived from (5.1). It is only possible when the environm

- **Choi:** a weird, but super fancy way, of representing channels. ent is very large.

We will investigate each of these ideas throughout this chapter. In the remainder of this section, I want to show you first how the four names appear in a specific example.

Kraus: amplitude damping

Consider a qubit system and let

$$M_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix}, \qquad M_1 = \begin{pmatrix} 0 & \sqrt{\lambda} \\ 0 & 0 \end{pmatrix}, \tag{5.4}$$

with $\lambda \in [0,1]$. This is a valid set of Kraus operators since $M_0^\dagger M_0 + M_1^\dagger M_1 = 1$. Its action on a general qubit density matrix reads:

$$\rho = \begin{pmatrix} p & q \\ q^* & 1-p \end{pmatrix} \qquad \to \qquad \rho' = \begin{pmatrix} \lambda + p(1-\lambda) & q\sqrt{1-\lambda} \\ q^*\sqrt{1-\lambda} & (1-\lambda)(1-p) \end{pmatrix}. \tag{5.5}$$

If $\lambda = 0$ nothing happens, $\rho' = \rho$. Conversely, if $\lambda = 1$ then

$$\rho \to \rho' = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{5.6}$$

This is why this is called an **amplitude damping**: no matter where you start, the map tries to push the system towards $|0\rangle$. It does so by destroying coherences,

$$q \to q\sqrt{1-\lambda}, \tag{5.7}$$

and by damping the populations,

$$p \to \lambda + p(1-\lambda). \tag{5.8}$$

The larger the value of λ, the stronger is the effect. This is **Kraus**.

Stinespring: system-environment interaction

Now let's look at **Stinespring**. Consider two qubits, which we label S (for system) and E (for environment). And assume that they interact unitarily with a Hamiltonian of the form

$$H = g(\sigma_+^S \sigma_-^E + \sigma_-^S \sigma_+^E), \tag{5.9}$$

where g is some parameter. The corresponding unitary evolution matrix will be

$$U = e^{-iHt} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos gt & -i\sin gt & 0 \\ 0 & -i\sin gt & \cos gt & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.10}$$

Now suppose that the ancila starts in the state $|0\rangle_E$ whereas the system starts in an arbitrary state ρ_S. They then evolve unitarily as in Eq. (5.1):

$$\rho'_{SE} = U\Big[\rho_S \otimes |0\rangle\langle 0|_E\Big]U^\dagger. \tag{5.11}$$

Taking the trace over E gives us the reduced density matrix of S:

$$\rho'_S = \mathrm{tr}_E\, \rho'_{SE} = \begin{pmatrix} p + (1-p)\sin^2(gt) & q\cos(gt) \\ q^*\cos(gt) & (1-p)\cos^2(gt) \end{pmatrix}. \tag{5.12}$$

But now comes the key point. If we look at this result for a second, we realize that it has the same structure as the amplitude damping result (5.5), provided we define

$$\lambda = \sin^2(gt). \tag{5.13}$$

Notice how subtle this is: I am not talking about the evolution in time. In fact, in the Kraus map (5.5) we don't even have time. What I am talking about is the structure of the *map*; the structure of how an input state ρ_S is processed into an output state ρ'_S. This is **Stinespring**: the effect of a quantum channel can be viewed as the interaction of a system with an environment, followed by a partial trace over the environment.

Lindblad: quantum master equation

Next consider a differential equation of the following form:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \gamma\Big[\sigma_+\rho\sigma_- - \frac{1}{2}\{\sigma_+\sigma_-, \rho\}\Big], \tag{5.14}$$

where $\{A, B\} = AB + BA$ is the anti-commutator. This is called a **Lindblad master equation**. One thing we can easily check is that it is trace-preserving: since the trace is linear, it follows from the cyclic property that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{tr}\,\rho = \gamma\,\mathrm{tr}\Big[\sigma_+\rho\sigma_- - \frac{1}{2}\{\sigma_+\sigma_-, \rho\}\Big] = 0. \tag{5.15}$$

Thus, $\mathrm{tr}\,\rho$ does not change in time. If it starts at 1, it stays at 1. Must less obvious is the fact that Eq. (5.14) preserves positivity. This is, in fact, why Lindblad got his name associated to this equation:[1] he showed that terms of the form $L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}$ always preserve positivity, provided the constant in front is positive.

Using a parametrization of the form (5.5), the master equation (5.14) can be split into two decoupled differential equations:

$$\begin{aligned} \frac{\mathrm{d}p}{\mathrm{d}t} &= \gamma(1-p) \qquad \to \qquad p(t) = p_0 e^{-\gamma t} + (1 - e^{-\gamma t}), \\ \frac{\mathrm{d}q}{\mathrm{d}t} &= -\frac{\gamma q}{2} \qquad \to \qquad q(t) = q_0 e^{-\gamma t/2}. \end{aligned}$$

---

[1] Around the same time Gorini, Kossakowski and Sudarshan also showed similar results. But, as with many things in science, Lindblad got most of the credit. That is slowly changing and a lot of people are now calling Eq. (5.14) the GKSL equation. Maybe I should do the same!

Comparing this with Eq. (5.5) we see that the evolution stemming from (5.14) can also be viewed as an amplitude damping channel, provided we identify

$$\lambda = 1 - e^{-\gamma t}. \tag{5.16}$$

If $t = 0$ then $\lambda = 0$ and nothing happens. If $t \to \infty$ then $\lambda \to 1$ and the system collapses completely towards $|0\rangle$, as in Eq. (5.6).

The names of Kraus, Stinespring and Lindblad are by far the most important and we will begin discussing them in more detail starting next section. But before doing so, I would like to tell you a bit about 3 slightly more advanced ideas. My goal is simply to illustrate how general and powerful the formalism we are dealing with is. If the above sections seem confusing, please skip to Sec. **??**.

Choi: weird sorcery

First, I want to tell you briefly about the Choi matrix, which is a powerful mathematical tool for proving theorems about quantum operations. Suppose that our system is actually composed of two qubits, A and B. We then act on only one of the two, say B, with the amplitude damping channel (5.4). This means that our quantum operation (5.2) will have the form

$$\rho'_{AB} = \mathcal{E}_B(\rho_{AB}) = \sum_k (I_A \otimes M_k)\rho_{AB}(I_A \otimes M_k^\dagger). \tag{5.17}$$

Here I wrote $\mathcal{E}_B$ to emphasize that the map is acting only on qubit B. Now consider the action of this map on a unnormalized Bell state,

$$|\Omega\rangle_{AB} = |00\rangle + |11\rangle. \tag{5.18}$$

As a result we get a matrix,

$$\Lambda_{\mathcal{E}} = \mathcal{E}_B(|\Omega\rangle\langle\Omega|_{AB}) = \begin{pmatrix} 1 & 0 & 0 & \sqrt{1-\lambda} \\ 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ \sqrt{1-\lambda} & 0 & 0 & 1-\lambda \end{pmatrix}. \tag{5.19}$$

This is the **Choi matrix**. The reason why it is useful is because, surprisingly, it turns out that it *completely characterizes the channel*. Every channel $\mathcal{E}$ has, associated to it, a corresponding Choi matrix $\Lambda_{\mathcal{E}}$. And the Choi matrix completely defines the channel in the sense that from $\Lambda_{\mathcal{E}}$ one can reconstruct $\mathcal{E}$. This is surprising because it means that, in order to know how a channel will act on *all* quantum states, it suffices to know how it acts on the maximally entangled state $|\Omega\rangle$ in Eq. (5.18). I know this is not at all obvious just by looking at Eq. (5.19). But I will show you next section how to do this.

Measurements and trace non-increasing maps

We saw in Sec. 2.7 that a generalized measurement was also defined by a set of Kraus operators $\{M_k\}$, where k labels the different outcomes of the measurement. And if outcome k is obtained, the state ρ is updated to

$$\rho \to \rho_k = \frac{M_k \rho M_k^\dagger}{p_k}, \tag{5.20}$$

where $p_k = \mathrm{tr}(M_k \rho M_k^\dagger)$. This map definitely takes valid density matrices to valid density matrices. However, it has a fundamental difference with respect to the maps we discussed before: it is *non-linear* because ρ also appears in the probabilities p_k. I think this kind of funny: professors and books always pound in our heads the idea that quantum mechanics is linear. But measurements are *not* linear.

We can recover linearity if we consider instead a map of the form

$$\mathcal{E}_k(\rho) = M_k \rho M_k^\dagger. \tag{5.21}$$

This is definitely linear, but it no longer preserves the trace. It is an example of a Completely Positive Trace Non-Increasing (CPTNI) map. The trace cannot increase because, since $\sum_k M_k^\dagger M_k = I$, it follows that $M_k^\dagger M_k \leq I$. The fact that it no longer preserves the trace may at first seem disturbing, but we can always normalize it whenever we need it, by writing $\rho_k = \mathcal{E}_k(\rho)/\,\mathrm{tr}\,\mathcal{E}_k(\rho)$. Thus, we can keep linearity, as long as we get used to working with states that are no longer normalized. That is the idea of "trace non-increasing maps".

Maps from one Hilbert space to the other

Finally, another powerful idea that I want to introduce (which will either make your head go boom, or make you extremely confused) is the notion that quantum channels can also be defined to map states from one system to another. Let us go back to the Stinespring system-environment map in Eq. (5.11). We calculated ρ'_{SE} and then traced over E to get ρ'_S; this formed a map that took us from ρ_S (a state of S) to ρ'_S (another state of S). But we could have, just as well, taken the trace over S. This would then lead to a state ρ'_E, which reads

$$\rho'_E = \mathrm{tr}_S\, \rho'_{SE} = \begin{pmatrix} p + (1-p)\cos^2(gt) & q\sin(gt) \\ q^* \sin(gt) & (1-p)\sin^2(gt) \end{pmatrix}. \tag{5.22}$$

This is similar to ρ'_S in Eq. (5.12), but with sines replaced by cosines and vice-versa. The point is that this result can now be viewed as a map from ρ_S to ρ_E:

$$\rho'_E = \mathcal{E}_{S\to E}(\rho_S). \tag{5.23}$$

This is a linear map, taking each state ρ_S of S to a state ρ'_E of E. Of course, this is for a given fixed initial state of E. If we were to use different ρ_E's, we would get different results. But in the spirit of *linear maps*, this is taking a state of S to a state of E. Here S and E have the same dimensions, but this does not have to be the case. One can also define maps form a Hilbert space of dimension d_1 to a Hilbert space of dimension d_2.

These kinds of maps are important because they represent the *transfer* of quantum states. The idea is that E may represent an empty register, $|0\rangle$. Then we come in with S and we wish to transfer the quantum state of S to the empty register E. The map (5.23) describes this process.

5.2 Stinespring representation theorem

In this section we make the Stinespring argument more rigorous. Consider a system S with arbitrary density matrix ρ_S. Now suppose this system is coupled to an environment E, which can have any dimension, but which we assume is prepared in a certain pure state $|\psi\rangle$ (we will lift this restriction in a second). The composite $S+E$ system then interacts via a unitary U (Fig. 5.1) after which we trace over the environment. This leads to a map

$$\boxed{\mathcal{E}(\rho_S) = \mathrm{tr}_E\Big\{U(\rho_S \otimes |\psi\rangle\langle\psi|_E)U^\dagger\Big\}.} \tag{5.24}$$

This map is clearly linear in ρ. It is also CPTP because we know that both unitary dynamics and partial traces are CPTP, so the operations we are doing are certainly on the safe side. In order to take the trace over E, we introduce a basis $|k\rangle_E$ for the environment, leading to

$$\begin{aligned}\mathcal{E}(\rho_S) &= \sum_k \langle k|U(\rho_S|\psi\rangle\langle\psi|)U^\dagger|k\rangle\\ &= \sum_k {}_E\langle k|U|\psi\rangle_E\, \rho_S\, {}_E\langle\psi|U^\dagger|k\rangle_E.\end{aligned} \tag{5.25}$$

This last step is a bit confusing, I know. What I did was split $|\psi\rangle\langle\psi|_E$ in two and pass the left one through ρ_S. I am allowed to do this because $|\psi\rangle_E$ lives in the space of the environment and thus commutes with ρ_S.[2]

The quantities ${}_E\langle k|U|\psi\rangle_E$ are still operators on the side of the system, as the contraction was done only on the side of the environment. If we define

$$\boxed{M_k = {}_E\langle k|U|\psi\rangle_E,} \tag{5.26}$$

we then see that Eq. (5.25) may be written in the form of a quantum channel,

$$\mathcal{E}(\rho_S) = \sum_k M_k \rho_S M_k^\dagger.$$

---

[2] You can also see this in a more clumsy way using tensor products. Any unitary U of SE can always be written as $U = \sum_\alpha A_\alpha \otimes B_\alpha$, where A, B are operators of S and E respectively. In tensor product notation, the kets $|k\rangle_E$ should be written as $I_S \otimes |k\rangle_E$. The map (5.24) may then be written as

$$\begin{aligned}\mathcal{E}(\rho_S) &= \sum_k \Big(I_S \otimes \langle k|\Big)\Big(\sum_\alpha A_\alpha \otimes B_\alpha\Big)\Big(\rho_S \otimes |\psi\rangle\langle\psi|\Big)\Big(\sum_\beta A_\beta^\dagger \otimes B_\beta^\dagger\Big)\Big(I_S \otimes |k\rangle\Big)\\ &= \sum_{\alpha,\beta} A_\alpha \rho_S A_\beta^\dagger\, \langle k|B_\alpha|\psi\rangle\langle\psi|B_\beta^\dagger|k\rangle\\ &= \Big(\sum_\alpha A_\alpha \langle k|B_\alpha|\psi\rangle\Big)\rho_S\Big(\sum_\beta A_\beta^\dagger\langle\psi|B_\beta^\dagger|k\rangle\Big),\end{aligned}$$

which is the same as Eq. (5.25).

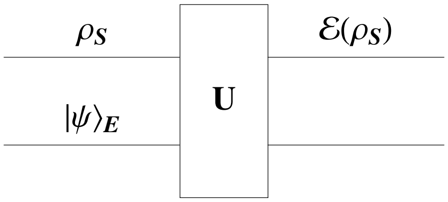

Figure 5.1: Idea behind a Stinespring dilation: a quantum operation $\mathcal{E}(\rho)$ can always be constructed by evolving the system together with an environment, with a global unitary U, and then discarding the environment.

We can also neatly see how the unitarity of U implies that the $\{M_k\}$ are indeed Kraus operators:

$$\sum_k M_k^\dagger M_k = \sum_k {}_E\langle\psi|U^\dagger|k\rangle_E \;\; {}_E\langle k|U|\psi\rangle_E = {}_E\langle\psi|U^\dagger U|\psi\rangle_E = I_S, \tag{5.27}$$

where I also used a completeness relation to eliminate the sum in k.

What we have just shown is that *whenever a system interacts unitarily with an environment, the dynamics from the point of view of the system, is given by a quantum operation*. Eq. (5.26) provides the ultimate link, expressing the Kraus operators in terms of (i) the system-environment interaction U and (ii) the environment initial state $|\psi\rangle$. These results are extremely powerful and form the basis for the entire theory of open quantum systems.

Generalization to mixed states of the environment

The calculation that led us to Eq. (5.26) assumed the environment started in a pure state $|\psi\rangle$. The generalization to a mixed initial state ρ_E is straightforward. We first diagonalize it as $\rho_E = \sum\limits_\mu q_\mu|\mu\rangle\langle\mu_E|$. Instead of Eq. (5.24), we then get

$$\begin{aligned}\mathcal{E}(\rho_S) &= \text{tr}_E\left\{U(\rho_S\otimes\rho_E)U^\dagger\right\}\\ &= \sum_{k,\mu} q_\mu \, {}_E\langle k|U|\mu\rangle_E \, \rho_S \, {}_E\langle\mu|U^\dagger|k\rangle_E.\end{aligned}$$

This has again the same structure as the Kraus decomposition, except that the Kraus operators now have two indices,

$$M_{k\mu} = \sqrt{q_\mu}\, {}_E\langle k|U|\mu\rangle_E. \tag{5.28}$$

The quantum channel then becomes

$$\mathcal{E}(\rho_S) = \sum_{k,\mu} M_{k\mu}\,\rho_S\, M_{k\mu}^\dagger. \tag{5.29}$$

Even though the two indices may seem like a complication, we can simply interpret (k, μ) as a collective index, so that formally this is not at all different from what we already had before. This also agrees with the notion of state purification, as discussed in Eq. (4.32): any mixed state ρ_E can always be viewed as a pure state in a larger Hilbert space. Hence, we could also simply apply the same recipe (5.26), but use instead a purified state $|\psi\rangle_{EE'} = \sum_\mu \sqrt{q_\mu}\, |\mu\rangle_E \otimes |\mu\rangle_{E'}$ living in a larger Hilbert space EE'.

Stinespring dilations

Eq. (5.26) provides a recipe on how to construct the set of Kraus operators given a U and a $|\psi\rangle$ of the environment. But what about the converse? Given a quantum channel, specified by a set of Kraus operators $\{M_k\}$, is it always possible to associate to it a certain system-environment unitary dynamics? The answer is yes. This is the idea of a **dilation**: any quantum channel can be represented as a unitary evolution in a dilated Hilbert space.

It can be shown that any quantum operation in a system with dimension d can always be described by an environment with dimension at most d^2. Thus, channels for a single qubit, for instance, can always be constructed by coupling it to an environment having 4 levels (e.g. two other qubits). The reason why this works is best shown using an example. Suppose we have the amplitude damping with Kraus operators (5.4). Now construct an environment in a state $|0\rangle$ and a unitary U of the form

$$U = \begin{pmatrix} 1 & 0 & x & x \\ 0 & \sqrt{1-\lambda} & x & x \\ 0 & \sqrt{\lambda} & x & x \\ 0 & 0 & x & x \end{pmatrix}. \tag{5.30}$$

Here x means arbitrary elements (not all equal) which should be included to ensure that U is indeed a unitary (the choices are not unique). Note how the first two columns are simply the entries of the Kraus operators (5.4) stacked together. Something like

$$U = \begin{pmatrix} M_0 & \dots \\ M_1 & \dots \end{pmatrix}.$$

This shows why the construction works: we simply build environments by hand, putting in U the elements of the M_k we want to have. It is, above all, mostly a matter of ordering elements.

Interpretation in terms of generalized measurements

Now that we know all about Stinespring, we can obtain a really neat interpretation for the generalized measurement postulate studied in Sec. 2.7: *A generalized measurement can always be viewed as the interaction between a system and an ancilla, plus a* *projective measurement in the ancilla.* Here is how this goes. Consider the $S + E$ interaction in Eq. (5.24). I am using the word ancilla now (which means auxiliary system) instead of environment, but it doesn't matter what you call it. After the interaction the global state of $S + E$ is

$$\rho'_{SE} = U(\rho_S \otimes |\psi\rangle\langle\psi|_E)U^\dagger,$$

We now perform a projective measurement in the ancilla, in the basis $|k\rangle$. This will update the state of the system to

$$\boxed{\rho_S^k = {}_E\langle k|U|\psi\rangle_E\, \rho_S\, {}_E\langle\psi|U^\dagger|k\rangle_E = M_k\, \rho_S\, M_k^\dagger,} \tag{5.31}$$

where I left out the normalization factor. Whence, we see that the generalized measurement $M_k \rho_S M_k^\dagger$ can always be viewed as the interaction of a system with an ancilla by means of some unitary U, followed by a projective measurement in the ancilla in the basis $|k\rangle$. The total quantum channel $\mathcal{E}(\rho)$ is then a sum of each possible outcome ρ_S^k (here we don't have to weight by the probabilities because the ρ_S^k are not normalized). The channel therefore encompasses all possible outcomes of the measurement. It is like measuring but not reading the outcomes of the measurement.

Unravelings and freedom in the operator-sum representation

What if we chose to measure the ancilla in a different basis, $|\alpha\rangle$? Well, of course, instead of Eq. (5.31) we would simply get

$$\rho_S^\alpha = {}_E\langle\alpha|U|\psi\rangle_E\, \rho_S\, {}_E\langle\psi|U^\dagger|\alpha\rangle_E. \tag{5.32}$$

If we now combine these states to form a quantum channel we get the same channel

$$\begin{aligned}\sum_\alpha \rho_S^\alpha &= \sum_\alpha {}_E\langle\alpha|U|\psi\rangle_E\, \rho_S\, {}_E\langle\psi|U^\dagger|\alpha\rangle_E \\ &= \text{tr}_E\Big\{U|\psi\rangle_E\, \rho_S\, {}_E\langle\psi|U^\dagger\Big\},\end{aligned}$$

where I simply used the fact that $\{|\alpha\rangle\}$ also forms a basis. Thus, we reach the important conclusion that as far as the channel is concerned, it doesn't matter which basis set you measure the ancilla. They all produce the same $\mathcal{E}(\rho)$.

However, depending on the choice of measurement basis, one may get different Kraus operators. These are called the different **unravelings** of the quantum channel. Indeed, Eq. (5.32), for instance, incites us to define

$$F_\alpha = {}_E\langle\alpha|U|\psi\rangle_E. \tag{5.33}$$

Hence,

$$\mathcal{E}(\rho_S) = \sum_k M_k\, \rho_S\, M_k^\dagger = \sum_\alpha F_\alpha\, \rho_S\, F_\alpha^\dagger. \tag{5.34}$$

The set of measurements $\{M_k\}$ and $\{F_\alpha\}$ represents different unravelings of the same channel. That is, different ways in which measurements can take place, but which lead to the same channel in the end. This ambiguity in the choice of Kraus operators is called **freedom in the operator-sum representation**. An "operator-sum representation" is a representation of the form $\sum_k M_k \rho_S M_k^\dagger$. A given channel $\mathcal{E}$, however, may have multiple operator sum representations.

The operators F_α in Eq. (5.33) and M_k in Eq. (5.26) are seen to be connected by a *basis transformation unitary* $V_{\alpha k} = \langle\alpha|k\rangle$:

$$_E\langle\alpha|U|\psi\rangle_E = \sum_k \langle\alpha|k\rangle\, _E\langle k|U|\psi\rangle_E$$

Whence,

$$\boxed{F_\alpha = \sum_k V_{\alpha k} M_k.} \tag{5.35}$$

In words: sets of Kraus operators connected by a unitary lead to the same quantum channel.

Example: qubit measurement

To illustrate these concepts consider a qubit prepared in an arbitrary state ρ_S and suppose we wish to measure the components in the σ_z basis. We couple this system to an ancilla prepared in $|0\rangle$, by means of a CNOT unitary

$$U = |0\rangle\langle 0|_S \otimes I_E + |1\rangle\langle 1|_S \otimes \sigma_x^E \tag{5.36}$$

If we measure the ancilla in the computational basis we then find the Kraus operators

$$M_0 = {_E\langle 0|U|0\rangle_E} = |0\rangle\langle 0|_S, \qquad M_1 = {_E\langle 1|U|0\rangle_E} = |1\rangle\langle 1|_S. \tag{5.37}$$

Thus, the outcomes we get from the ancilla represent the probabilities of finding the system in 0 and 1:

$$p_i = \operatorname{tr}\left\{M_i \rho_S M_i^\dagger\right\} = \langle i|\rho_S|i\rangle, \qquad i = 0, 1.$$

But now suppose we decide to measure the ancillas in a different basis, say $|\pm\rangle_E$. Then the Kraus operators become

$$F_0 = {_E\langle +|U|0\rangle_E} = \frac{I_S}{\sqrt{2}}, \qquad F_1 = {_E\langle -|U|0\rangle_E} = \frac{\sigma_z^S}{\sqrt{2}}. \tag{5.38}$$

The measurement outcomes are now completely uninformative, $q_0 = q_1 = 1/2$. Notwithstanding, the channel is still exactly the same.

5.3 Choi's matrix and proof of the Kraus representation

All I want to do in this section is to prove the following mathematical result: Let $\mathcal{E}(\rho)$ denote a map satisfying

1. Linearity: $\mathcal{E}(\alpha\rho_1 + \beta\rho_2) = \alpha\mathcal{E}(\rho_1) + \beta\mathcal{E}(\rho_2)$.
2. Trace preserving: $\text{tr}[\mathcal{E}(\rho)] = \text{tr}(\rho)$.
3. Completely positive: if $\rho \geq 0$ then $\mathcal{E}(\rho) \geq 0$.

Then this map can always be represented in the form (5.2) for a certain set of Kraus operators $\{M_k\}$. Proving this is nice because it is a way of showing that the quantum channel is *the* most general map satisfying 1,2 and 3.

There is a subtle difference between a map that is positive and a map that is *completely* positive. Completely positive means $\mathcal{E}(\rho) \geq 0$ even if ρ is a density matrix living in a larger space than the one $\mathcal{E}$ acts on. For instance, suppose $\mathcal{E}$ acts on the space of a qubit. But maybe we want to apply this map to the density matrix of 2 entangled qubits. If even in this case the resulting ρ' is positive semi-definite, we say the map is completely positive.[3]

The proof of our claim is based on a powerful, yet abstract, idea related to what is called the *Choi isomorphism*. Let S denote the space where our map $\mathcal{E}$ acts and define an auxiliary space R which is an exact copy of S. Define also the (unnormalized) Bell state

$$|\Omega\rangle = \sum_i |i\rangle_R \otimes |i\rangle_S, \tag{5.39}$$

where $|i\rangle$ is an arbitrary basis and from now on I will always write the R space in the left and the S space in the right. We now construct the following operator:

$$\boxed{\Lambda_{\mathcal{E}} = (\mathcal{I}_R \otimes \mathcal{E}_S)(|\Omega\rangle\langle\Omega|).} \tag{5.40}$$

This is called the **Choi matrix** of the map $\mathcal{E}$. It is the outcome of applying the map $\mathcal{E}_S$ on one side of the maximally entangled Bell state of R+S. Hence, it is a bit like a density matrix in the space RS: it is positive definite although not normalized.

The most surprising thing about the Choi matrix is that it completely determines the map $\mathcal{E}$. That is, if we somehow learn how our map $\mathcal{E}$ acts on $|\Omega\rangle\langle\Omega|$ we have completely determined how it will act on *any* other density matrix. This is summarized by the following formula:

$$\boxed{\mathcal{E}(\rho) = \text{tr}_R\Big\{(\rho^{\text{T}} \otimes I_S)\Lambda_{\mathcal{E}}\Big\}.} \tag{5.41}$$

---

[3] There aren't many examples of maps that are positive but not completely positive. The only example I know is the partial trace (see, for instance, Box 8.2 of Nielsen and Chuang).

I know what you are thinking that this is super weird. And I agree! It is! But now let's check and see that this sorcery actually works.

Note that here ρ^{T} is placed on the auxiliary space R in which the trace is being taken. Consequently, the result on the left-hand side is still an operator living on S. To verify that Eq. (5.41) is true we first rewrite (5.40) as

$$\Lambda_E = \sum_{i,j} |i\rangle_R\langle j| \otimes \mathcal{E}(|i\rangle\langle j|). \tag{5.42}$$

Then we get

$$\begin{aligned}
\mathrm{tr}_R\Big\{(\rho^{\mathrm{T}} \otimes I_S)\Lambda_{\mathcal{E}}\Big\} &= \sum_{i,j} \mathrm{tr}_R\Big\{(\rho^{\mathrm{T}} \otimes I_S)\Big[|i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|)\Big]\Big\} \\
&= \sum_{i,j} \langle j|\rho^{\mathrm{T}}|i\rangle \mathcal{E}(|i\rangle\langle j|) \\
&= \mathcal{E}\Big(\sum_{i,j} \rho_{i,j}|i\rangle\langle j|\Big) \\
&= \mathcal{E}(\rho).
\end{aligned}$$

Here I used the fact that $\langle j|\rho^{\mathrm{T}}|i\rangle = \langle i|\rho|j\rangle = \rho_{i,j}$. Moreover, I used our assumption that $\mathcal{E}$ is a linear map.

We are now in the position to prove our claim. As I mentioned, the Choi matrix looks like a density matrix on R+S. In fact, we are assuming that our map $\mathcal{E}$ is CPTP. Thus, since $|\Omega\rangle\langle\Omega|$ is a positive semi-definite operator, then so will $\Lambda_{\mathcal{E}}$ (although it will not be normalized). We may then diagonalize $\Lambda_{\mathcal{E}}$ as

$$\Lambda_{\mathcal{E}} = \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|,$$

where $|\lambda_k\rangle$ are vectors living in the big R+S space and $\lambda_k \geq 0$. For the purpose of what we are going to do next, it is convenient to absorb the eigenvalues into the eigenvectors (which will no longer be normalized) and define

$$\Lambda_{\mathcal{E}} = \sum_k |m_k\rangle\langle m_k|, \qquad |m_k\rangle = \sqrt{\lambda_k}|\lambda_k\rangle, \tag{5.43}$$

Note that here CPTP is crucial because it implies that $\lambda_k \geq 0$ so that $\langle m_k| = \langle\lambda_k|\sqrt{\lambda_k}$. To finish the proof we insert this into Eq. (5.41) to get

$$\mathcal{E}(\rho) = \sum_k \mathrm{tr}_R\Big\{(\rho^{\mathrm{T}} \otimes I_S)|m_k\rangle\langle m_k|\Big\}. \tag{5.44}$$

The right-hand side will still be an operator living in S, since we only traced over R. All we are left to do is convince ourselves that this will have the shape of the operator-sum representation in Eq. (5.2).

In order to do that, things will get a bit nasty. The trick is to connect the states $|m_k\rangle$ of the Choi matrix $\Lambda_{\mathcal{E}}$ with the Kraus operators M_k appearing in the operator-sum representation (5.2): This is done by noting that since $|m_k\rangle$ lives on the R+S space, it can be decomposed as

$$|m_k\rangle = \sum_{i,j} (M_k)_{j,i} |i\rangle_R \otimes |j\rangle_S, \tag{5.45}$$

where $(M_k)_{j,i}$ are a set of coefficients which we can interpret as a matrix M_k. We now manipulate (5.44) to read

$$\mathcal{E}(\rho) = \sum_k \sum_{i,j} {}_R\langle i|\rho^{\mathrm{T}}|j\rangle_R \, {}_R\langle j|m_k\rangle\langle m_k|i\rangle_R.$$

Then we insert Eq. (5.45) to find

$$\begin{aligned}\mathcal{E}(\rho) &= \sum_k \sum_{i,j} \sum_{i',j'} \rho_{j,i} (M_k)_{j',j} (M_k^*)_{i',i} |j'\rangle\langle i'| \\ &= \sum_k \sum_{i,j} \sum_{i',j'} |j'\rangle\langle j'|M_k|j\rangle\langle j|\rho|i\rangle\langle i|M_k^\dagger|i'\rangle\langle i'|. \\ &= \sum_k M_k \rho M_k^\dagger,\end{aligned}$$

and voilá!

In conclusion, we have seen that any map which is linear and CPTP can be described by an operator-sum representation, Eq. (5.2). I like this a lot because we are not asking for much: linearity and CPTP is just the *basic* things we expect from a physical map. Linearity should be there because everything in quantum mechanics is linear and CPTP must be there because the evolution must map a physical state into a physical state. When we first arrived at the idea of a unitary, we were also very relaxed because all we required was the conservation of ket probabilities. The spirit here is the same. For this reason, the quantum operation is really just a very natural and simplistic generalization of the evolution of quantum systems, using density matrices instead of kets.

5.4 Lindblad master equations

Quantum operations describe a map from the initial to the final state. They are very powerful on a conceptual level, but not always easy to apply, specially if your system is more complicated than 1 or 2 qubits. Conversely, sometimes we also want to have access to the *evolution*; that is, to $\rho(t)$. And we want to have a method which is easier to use and, perhaps, also easier to simulate in a computer. This is where Lindblad comes in.

We want to describe the evolution of a system coupled to an environment. If the system were isolated it would evolve according to von Neumann's equation,

$$\frac{d\rho}{dt} = -i[H, \rho].$$

But when there is also the coupling to an environment, we expect an evolution equation of the form

$$\boxed{\frac{\mathrm{d}\rho}{\mathrm{d}t} = -i[H,\rho] + D(\rho),} \tag{5.46}$$

where $D(\rho)$ is an additional term, called the **dissipator**, which describes the effects of the environment. This type of equation is historically known as a **master equation**, a name which was first introduced in a completely different problem,[4] and is supposed to mean an equation from which all other properties can be derived from. I think its a weird name. Sounds like a Dan Brown book: *the master equation*.

In this section I want to address the question of what are the typical forms one can have for $D(\rho)$ that lead to a valid CPTP evolution. That is, such that the solution $\rho(t)$ of Eq. (5.46) is a positive semidefinite density matrix at all times t. This is the content of *Lindblad's theorem*, who showed that in order for this to happen, $D(\rho)$ must have the form

$$\boxed{D(\rho) = \sum_k \gamma_k \Big[L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}\Big], \qquad \gamma_k \geq 0,} \tag{5.47}$$

where L_k are arbitrary operators. If you have any equation satisfying this structure, then the corresponding evolution is guaranteed to be CPTP (i.e., physical). Master equations having this structure are then called Linbdlad master equations or, more generally, Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) equations.

The operator $D(\rho)$ in Eq. (5.46) is called a **superoperator**. It is still a linear operator, as we are used to in quantum mechanics. But it acts on density matrices, instead of kets, which means it can multiply ρ on both sides. Notwithstanding, it is essential to realize that despite this complication, Eq. (5.46) still has the general structure of a linear equation

$$\boxed{\frac{\mathrm{d}\rho}{\mathrm{d}t} = \mathcal{L}(\rho),} \tag{5.48}$$

The superoperator $\mathcal{L}(\rho)$ is called the **Liouvillian** (because of the analogy Liouville's equation in classical mechanics). This equation is just like a matrix vector equation

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = A\boldsymbol{x}, \tag{5.49}$$

where $\boldsymbol{x}$ is a vector and A is a matrix. The density matrix is now the "vector" (matrices form a vector space) and the Liouvillian $\mathcal{L}$ is now the "matrix" (linear operator acting on the vector space).

The solution of Eq. (5.49) is well known. It is simply $\boldsymbol{x}(t) = e^{At}\boldsymbol{x}(0)$, which is the *map* from the initial to the final state. We can call $B = e^{At}$ as the **map**. Then A is the

---

[4] A. Nordsieck, W. E. Lamb and G. T. Uhlenbeck, *Physica*, **7**, 344 (1940).

generator of the map. The same will be true for Eq. (5.48); that is, we can also write

$$\rho(t) = \mathcal{E}_t(\rho(0)), \tag{5.50}$$

where $\mathcal{E}_t$ is the linear map taking $\rho(0)$ to $\rho(t)$, which is related to the *generator* $\mathcal{L}$ by means of $\mathcal{E}_t = e^{\mathcal{L}t}$.[5]

The letter $\mathcal{E}$ already hints at where I want to get to eventually: *If the generator is in Lindblad form (5.47), then the map will be a quantum channel (5.2).* That is the essence of Lindblad's theorem. But what about the converse? When can a quantum channel be expressed as the exponential of a Lindblad generator?

To answer this requires us to introduce another property of equations of the form (5.48), known as the **semigroup property**. Using again the analogy with the linear Eq. (5.49), we know that we can split the evolution into multiple steps. If we first evolve to t_1 and then evolve for an extra t_2, it is the same as if we evolved all the way through by $t_1 + t_2$. The matrix exponential makes this obvious: $e^{At_2}e^{At_1} = e^{A(t_2+t_1)}$. Since the master equation has the same structure, this must also be true for the map $\mathcal{E}_t$. That is, it must satisfy the semigroup property [6]

$$\mathcal{E}_{t_2}\mathcal{E}_{t_1} = \mathcal{E}_{t_1+t_2}. \tag{5.51}$$

Semigroup is therefore implied by the structure of Eq. (5.48).

Lindblad's Theorem

We can now update our question: *What is the structure of a map which is both CPTP and semigroup?* This is the content of **Lindblad's theorem**:[a] The generator of any quantum operation satisfying the semigroup property must have the form:

$$\boxed{\frac{ud\rho}{\mathrm{d}t} = \mathcal{L}(\rho) = -i[H,\rho] + \sum_k \gamma_k\Big[L_k\rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k,\rho\}\Big],} \tag{5.52}$$

where H is a Hermitian operator, L_k are arbitrary operators and $\gamma_k \geq 0$. If you have any equation satisfying this structure, then the corresponding evolution is guaranteed to be CPTP (i.e., physical). Conversely, any CPTP and divisible map is guaranteed to have to this form.

[a] G. Lindblad, *Comm. Math. Phys*, **48**, 119 (1976).

---

[5] This formula is pretty on a formal level, but it is difficult to apply because $\mathcal{L}$ does not multiply ρ on the left only. But luckily we won't have to enter into this issue.

[6] Here "group" refers to the group of CPTP maps $\mathcal{E}_t$ characterized by a single parameter t. Eq. (5.51) is the composition property for a group. But the reason why it is a *semi*group, is because but the inverse is not a member of the group (as required for a set to be called a group). Here the inverse is $\mathcal{E}_{-t}$. While this exists, it is not in general CPTP (unless $\mathcal{E}$ is unitary, as Lindblad shows).

Of course, this does not say anything about how to *derive* such an equation. That is a hard question, which we will start to tackle in the next section. But this result gives us an idea of what kind of structure we should look for and that is already remarkably useful.

Proof of Lindblad's theorem

If the dynamics is to satisfy the semigroup property (i.e., if it is divisible) then we must be able to write the evolution over an infinitesimal time Δt as

$$\rho(t+\Delta t) = \sum_k M_k(\Delta t)\rho(t)M_k^\dagger(\Delta t), \tag{5.53}$$

where the Kraus operators $M_k(\Delta t)$ cannot depend on the time t. We are interested in a differential equation for $\rho(t)$, which would look something like

$$\rho(t+\Delta t) \simeq \rho(t) + \Delta t \mathcal{L}(\rho(t)).$$

Since the first correction is of the order Δt, we then expect to have $M_k(\Delta t) = \sqrt{\Delta t} L_k$, where L_k is some operator. This is so because then $M_k \rho M_k^\dagger \sim \Delta t$. But we also have the additional property that, if $\Delta t = 0$, then nothing should happen: $\sum_k M_k(0)\rho M_k(0) = \rho$. So not all M_k can be of order $\sqrt{\Delta t}$. At least one must be of order 1.

Let us try a parametrization of the following form

$$M_0 = I + G\Delta t, \qquad\qquad M_k = \sqrt{\gamma_k \Delta t} L_k, \qquad k \neq 0$$

where G and L_k are arbitrary operators and $\gamma_k \geq 0$ are constants that I introduce simply to make the L_k dimensionless. The normalization condition for the Kraus operators implies that

$$\begin{aligned} I = \sum_k M_k^\dagger M_k &= M_0^\dagger M_0 + \sum_{k\neq 0} M_k^\dagger M_k \\ &= (I + G^\dagger \Delta t)(I + G\Delta t) + \Delta t \sum_{k\neq 0} \gamma_k L_k^\dagger L_k \\ &= I + (G + G^\dagger)\Delta t + \Delta t \sum_{k\neq 0} \gamma_k L_k^\dagger L_k + \mathbb{O}(\Delta t^2). \end{aligned}$$

This shows why we need this G guy. Otherwise, we would never be able to normalize the Kraus operators. Since G is arbitrary, we may parametrize it as

$$G = K - iH, \tag{5.54}$$

where K and H are both Hermitian. It then follows from the normalization condition that

$$K = -\frac{1}{2}\sum_{k\neq 0} \gamma_k L_k^\dagger L_k, \tag{5.55}$$

whereas nothing can be said about H. The operator G then becomes

$$G = -\Big\{iH + \frac{1}{2}\sum_{k\neq 0}\gamma_k L_k^\dagger L_k\Big\}. \tag{5.56}$$

This concludes our recipe for constructing M_k and M_0. They are now properly normalized to order Δt. And when $\Delta t \to 0$ we get only $M_0 = I$ (nothing happens).

With this at hand, we can finally substitute our results in Eq. (5.53) to find

$$\begin{aligned}
\rho(t+\Delta t) &= (I + G\Delta t)\rho(I + G^\dagger \Delta t) + \Delta t\sum_{k\neq 0}\gamma_k L_k\rho L_k^\dagger \\
&= \rho(t) + \Delta t(G\rho + \rho G^\dagger) + \Delta t \sum_{k\neq 0}\gamma_k L_k \rho L_k^\dagger \\
&= \rho(t) - i\Delta t[H,\rho] + \Delta t\sum_{k\neq 0}\gamma_k\Big[L_k\rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}\Big]
\end{aligned}$$

Rearranging and taking the limit $\Delta t \to 0$ we then finally obtain

$$\frac{\rho(t+\Delta t) - \rho(t)}{\Delta t} \simeq \frac{\mathrm{d}\rho}{\mathrm{d}t} = -i[H,\rho] + \sum_{k\neq 0}\gamma_k\Big[L_k\rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k,\rho\}\Big], \tag{5.57}$$

which is Lindblad's equation (5.52). Woooooo-hoooooo!!! We did it! We found that if we combine the semigroup property and the structure of a quantum operation, the corresponding differential equation must have Lindblad's form. As I mentioned before, we still have no idea of what the operators H and L_k should be. That will be the topic of next section. But it is great that we can already tell what the general structure should be.

Intuition behind Lindblad jump operators

The calculations above established the connection between Lindblad and Kraus, which can be summarized by

$$M_0 = I - \Delta t\Big\{iH + \frac{1}{2}\sum_{k\neq 0}\gamma_k L_k^\dagger L_k\Big\}, \qquad M_k = \sqrt{\gamma_k \Delta t}L_k, \qquad k\neq 0. \tag{5.58}$$

These equations can also be used in the reverse way: it tells you what will be the corresponding Kraus operators if you integrate the Lindblad Eq. (5.52) over an infinitesimal time Δt. This is neat because we already have a nice intuition for the action of the Kraus operators as generalized measurements. We know that the M_k cause *quantum jumps*: abrupt transitions in the state of the system as $\rho \to M_k\rho M_k^\dagger$. Over a small time interval Δt, there is a large probability that nothing happens (M_0) and a small probability that one of the jumps M_k occur. The operators L_k are called **jump operators**. They represents the types of jumps one may have in the quantum dynamics.

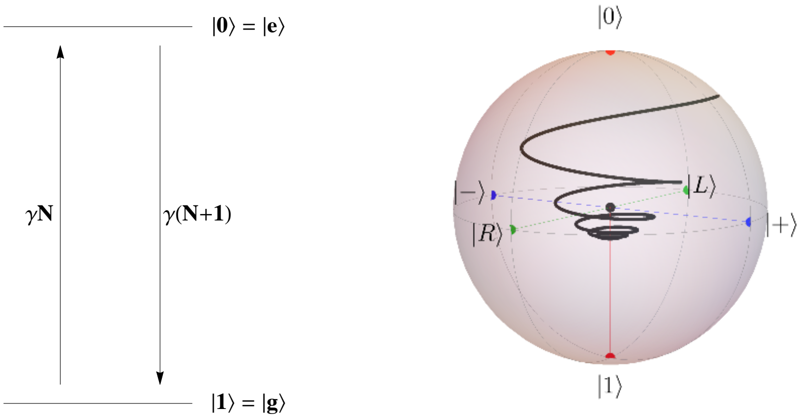

Figure 5.2: Example evolution of $\rho(t)$ under the map (5.59). **Left:** energy level diagram showing also the transition rates. **Right:** dynamics in the Bloch sphere. The initial state is taken as $|\psi\rangle = (\cos\frac{\pi}{8}, e^{i\pi/4}\sin\frac{\pi}{8})$ and the parameters were

Example: Finite temperature amplitude damping

Consider a two-level atom with Hamiltonian $H = \frac{\Omega}{2}\sigma_z$. The ground-state is then the south pole $|1\rangle$ in the Bloch sphere. The contact of this atom with the surrounding electromagnetic field will produce an evolution described by the master equation

$$\frac{d\rho}{dt} = -i[H,\rho] + \gamma(N+1)D[\sigma_-] + \gamma N D[\sigma_+], \tag{5.59}$$

where γ is a constant and

$$N = \frac{1}{e^{\beta\Omega} - 1},$$

is the Bose-Einstein distribution with inverse temperature $\beta = 1/k_B T$ and frequency Ω. In Eq. (5.59) I also introduce the notation

$$D[L] = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}. \tag{5.60}$$

After all, since the dissipator is fully specified by the jump operator L, we don't need to write the full dissipator all the time.

An example of the evolution of the density matrix under Eq. (5.59) is shown in Fig. 5.2. The Hamiltonian part induces the qubit to precess around the z axis. If there was no dissipation the spin would precess indefinitely. But in the presence of dissipation, it precesses and is also damped towards the z axis.

Steady-state: After a long time has elapsed the system will tend to a steady-state, which is the solution of

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \mathcal{L}(\rho) = 0. \tag{5.61}$$

You may check that the steady-state is in this case diagonal, with

$$\langle \sigma_x \rangle_{\mathrm{ss}} = \langle \sigma_y \rangle_{\mathrm{ss}} = 0, \qquad \langle \sigma_z \rangle_{\mathrm{ss}} = -\frac{1}{2N+1}, \tag{5.62}$$

which corresponds to a thermal equilibrium density matrix

$$\rho_{\mathrm{ss}} = \begin{pmatrix} \frac{N}{2N+1} & 0 \\ 0 & \frac{N+1}{2N+1} \end{pmatrix}. \tag{5.63}$$

Emission and absorption: Let us now understand this from the perspective of the jump operators (5.58). Our master equation (5.59) is characterized by jump operators σ_- and σ_+, which lower and raise the energy of the atom respectively. Moreover, these transitions occur with rates $\gamma(N+1)$ and γN, so that it is more likely to observe a transition downwards than upwards. A transition upwards represents the absorption of a photon and a transition downwards represents an emission. In the limit of zero temperature we get $N \to 0$. In this case the atom only interacts with the electromagnetic vacuum. As a consequence, the absorption tends to zero. However, the emission does not. This is the idea behind **spontaneous emission**. Even in an environment with zero photons on average, the atom still interacts with the vacuum and can emit radiation. This is the physical meaning of the factor "1" in $N+1$.

Unitary vs. dissipative: The dissipative dynamics of Eq. (5.59) is described by the jump operators σ_- and σ_+. We could also have introduced jump operators directly in the Hamiltonian. For instance, consider the *purely unitary* dynamics under

$$H = \frac{\Omega}{2}\sigma_z + \frac{\lambda}{2}(\sigma_+ + \sigma_-). \tag{5.64}$$

This type of Hamiltonian appears when the atom is pumped by a coherent light source (i.e. a laser). How similar is this dynamics from the Lindblad equation? They are *completely different*. First, note that Hamiltonians have to be Hermitian, so the rates for upward and downward transitions have to be equal. This reflects the fact that unitary dynamics is always **reversible**. An example of the dynamics generated by Eq. (5.64) is shown in Fig. 5.3. As can be seen, all it does is shift the axis of precession. Instead of precessing around z, it will precess around some other axis. There is no damping. In unitary dynamics there is *never* any damping.

Dissipators and Hamiltonians compete: Let us now suppose, for some reason, that the system evolves according to the MEq. (5.59), but with the Hamiltonian (5.64). This could mean, for instance, a damped atom also pumped by an external laser. We want to compute the steady-state ρ_{ss}, which is the solution of

$$-i[\frac{\Omega}{2}\sigma_z + \frac{\lambda}{2}(\sigma_+ + \sigma_-), \rho_{\mathrm{ss}}] + \gamma(N+1)D[\sigma_-] + \gamma N D[\sigma_+] = 0. \tag{5.65}$$

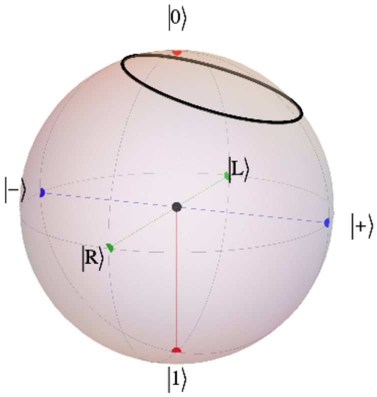

Figure 5.3: Unitary evolution of $\rho(t)$ under the Hamiltonian (5.64). Compare this with the dissipative dynamics in Fig. 5.2. The term σ_x also causes transitions, but the Hamiltonian dynamics is fundamentally different from the open dynamics.

The point I want to emphasize about this equation, is that the steady-state depends on both the unitary as well as the dissipative part. The steady-state (5.63) happened not to depend on the Hamiltonian which was simply σ_z. But in general, dissipators and Hamiltonians compete. And the steady-state is somewhere in the middle. In the case of Eq. (5.65) the steady-state turns out to be quite ugly, and is given by

$$\langle\sigma_x\rangle_{\text{ss}} = -\frac{4\lambda\Omega}{(2N+1)[\gamma^2(2N+1)^2 + 2(\lambda^2 + 2\Omega^2)]}, \tag{5.66}$$

$$\langle\sigma_y\rangle_{\text{ss}} = -\frac{2\gamma\lambda}{[\gamma^2(2N+1)^2 + 2(\lambda^2 + 2\Omega^2)]}, \tag{5.67}$$

$$\langle\sigma_z\rangle_{\text{ss}} = -\frac{4\Omega^2 + \gamma^2(2N+1)^2}{(2N+1)[\gamma^2(2N+1)^2 + 2(\lambda^2 + 2\Omega^2)]}. \tag{5.68}$$

In addition to ugly, this steady-state is also not very intuitive. For instance, if we take the limit $\lambda \to 0$ we recover the results in (5.62). But if we instead take the limit $\lambda \to \infty$ (a very strong laser) we get instead the maximally mixed state $\langle\sigma_i\rangle = 0$. The message I want you to take from this is that the steady-state of a master equation is always a competition between different terms. Moreover, Hamiltonian and dissipative terms compete in different ways. As a consequence, the steady-state is not always intuitive.

5.5 Collisional models

In this section we get into the issue on how to *derive* a Lindblad equation. The first thing one must bear in mind is that these derivations are *not* universal; they are model specific, depending not only on the system, but also on the type of environment *and* the specific system-environment interaction. Hence, they should be derived on a case-by-case basis. Of course, with some practice, one can start to gain some intuition as to what the Lindblad equations should look like, and so on.

So to start this adventure, I propose we study a very simple model of master equations, which I absolutely love. They are called collisional models and are illustrated in Fig. 5.4. A system S, prepared in an arbitrary state ρ_S is allowed to interact with a sequence of ancillas. The ancillas are all independent and have been prepared in identical states ρ_E (each ancilla can be a qubit, for instance, so ρ_E is the 2×2 matrix of a single qubit). Each interaction is described by some unitary $U_{SE}(\tau)$ (in the same spirit as what we did in Stinespring's theorem in Sec. 5.2) that lasts for some duration τ. Thus, if we think about one system-ancilla interaction event, their global state at the end will be given by

$$\rho_{SE'} = U(\rho_S \otimes \rho_E)U^\dagger, \tag{5.69}$$

where, in order to make the notation more compact, I write only U instead of $U_{SE}(\tau)$.

After one system-ancilla interaction, we then throw away the ancilla and bring in a fresh new one, again prepared at the state ρ_E. Since we threw away the ancilla, the state of the system is now $\rho'_S = \operatorname{tr}_E \rho'_{SE}$. We then repeat the process, using ρ'_S as the input in Eq. (5.69) and evolving the system towards another state ρ''_S. All we care about is the **stroboscopic** evolution of the system, in integer multiples of τ. In fact, we can make this more precise as follows. Let $\rho_S(n) \equiv \rho_S(n\tau)$ denote the state of the system after n interactions. So the initial state is $\rho_S(0)$, the state after the first interaction is $\rho_S(1)$ and etc. The game of the collisional model is therefore given by the stroboscopic map

$$\boxed{\rho_S(n+1) = \operatorname{tr}_E \Big\{ U(\rho_S(n) \otimes \rho_E)U^\dagger \Big\}.} \tag{5.70}$$

This kind of evolution is really neat. It is exactly the map of the Stinespring dilation [Eq (5.24)], but applied multiple times, each with a brand new environment. This map is therefore *constructed* to satisfy the semigroup property at the stroboscopic level.

Lindblad master equation for qubit ancillas

The map (5.70) can be defined for an arbitrary system-ancilla interaction time τ, as it only depends on the unitary involved. To obtain a Lindblad equation, we now consider the scenario where the interaction times are extremely short. This will allow us to approximate the time derivative of ρ_S as

$$\frac{d\rho_S}{dt} \simeq \frac{\rho_S(n+1) - \rho(n)}{\tau}. \tag{5.71}$$

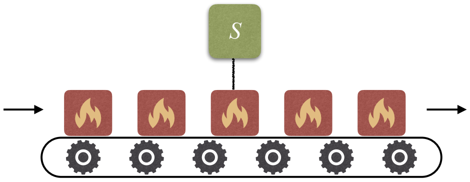

Figure 5.4: Illustration of a collisional model: a system S is allowed to interact sequentially with a set of ancillas, all prepared in the same state and all interactions lasting only for a short amount of time τ. After the interaction the ancilla is thrown away and a fresh new one is introduced, in the same state.

I'm not saying we take the "limit of $\tau \to 0$". There is no limit in the mathematical sense here: τ is still finite. But if it is sufficiently small, compared to other time scales, then the discrete difference $\rho_S(n+1) - \rho(n)$ could be a good approximation for a derivative.

In order to gain intuition, let us introduce a specific model for system-ancilla interactions. The system is assumed to be arbitrary, but the ancillas are taken to be qubits prepared in a diagonal (thermal) state

$$\rho_E = \begin{pmatrix} f & 0 \\ 0 & 1-f \end{pmatrix} = f|0\rangle\langle 0| + (1-f)|1\rangle\langle 1|, \tag{5.72}$$

where $f \in [0,1]$ is the probability of finding the ancilla in $|0\rangle$. We then assume that the unitary U is generated by a system ancilla interaction of the form

$$V = g(L^\dagger \sigma_-^E + L\sigma_+^E), \tag{5.73}$$

where g is a real number and L is an arbitrary operator of the system. This type of interaction is very common and is sometimes referred to as an exchange Hamiltonian: it essentially says that whenever the qubit goes up a level (by applying σ_+^E), one should apply L to the system (whatever the meaning of L is). Whereas if the qubit goes down a level, we apply $L^\dagger$. For instance, if our system was another qubit, maybe we could just use $L = \sigma_-^S$. Then if the bath goes up, the system goes down. But in general, the choice of L is arbitrary.

The unitary U in Eq. (5.70) will then be given by $U = e^{-i\tau V}$. For simplicity, I will not worry here about the Hamiltonians of the system and environment, just about the interaction V. We now consider the unitary map (5.69) and expand the sandwich of exponentials using the BCH formula:

$$\rho_{SE}' = e^{-i\tau V}\rho_S\rho_E e^{i\tau V} = \rho_S\rho_E - i\tau[V, \rho_S\rho_E] - \frac{\tau^2}{2}[V,[H_{SE}, \rho_S\rho_E]] + \dots \tag{5.74}$$

To obtain the equation for S, we then trace over the environment. The first term is trivial: $\text{tr}_E\, \rho_S\rho_E = \rho_S$. In the second term we have

$$\text{tr}_E[V, \rho_S\rho_E] = \text{tr}_E\Big(V\rho_S\rho_E - \rho_S\rho_E V\Big).$$

This is the trickiest term: you cannot use the cyclic property of the trace because now V acts on both subspaces of S and E, whereas the trace is only partial. But what we can do is commute ρ_S and ρ_E at will, so that this may be written as

$$\mathrm{tr}_E[V, \rho_S\rho_E] = \mathrm{tr}_E(V\rho_E)\rho_S - \rho_S\,\mathrm{tr}_E(\rho_E V).$$

However, due to our choice of state ρ_E and interaction V in Eqs. (5.72) and (5.73), we have that

$$\mathrm{tr}_E(\sigma_\pm^E\rho_E) = 0.$$

Hence, this term is also identically zero:

$$\mathrm{tr}_E[V, \rho_S\rho_E] = 0.$$

Hence, taking the partial trace of Eq. (5.74) yields

$$\rho_S' = \rho_S - \frac{\tau^2}{2}\,\mathrm{tr}_E[V, [V, \rho_S\rho_E]]. \tag{5.75}$$

Dividing by τ and arranging things so as to appear the derivative in Eq. (5.71), we find

$$\frac{\mathrm{d}\rho_S}{\mathrm{d}t} = D(\rho_S) := -\frac{\tau}{2}\,\mathrm{tr}_E[V, [V, \rho_S\rho_E]]. \tag{5.76}$$

This result is a bit weird, because there is still a τ in there. This is why I said above that we should not take the limit $\tau \to 0$. The approximation as a derivative in Eq. (5.71) does not mean that anything multiplying τ should be negligible. The reason is that V still has a constant g [Eq. (5.73)]. So it can be that even though τ is small, $g^2\tau$ will be finite.

This type of issue also appears in classical stochastic processes. For instance, the Langevin equation describing Brownian motion is

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\gamma\frac{\mathrm{d}x}{\mathrm{d}t} + f(x) + \sqrt{2\gamma k_B T}\,\xi(t),$$

where $\xi(t)$ is a random noise satisfying $\langle\xi(t)\xi(t')\rangle = \delta(t-t')$. The noise is a delta function, so that it acts only for an infinitesimal time interval. But if you want to get something non-trivial, it must also be infinitely large to compensate. The logic is the same in Eq. (5.76): we are assuming the interaction time τ with each ancilla is very short. But if we want to get something non-trivial out of it, we must also make this interaction sufficiently strong.

In any case, continuing with our endeavor, we now write

$$-[V, [V, \rho_S\rho_E]] = 2V\rho_S\rho_E V - V^2\rho_S\rho_E - \rho_S\rho_E V^2.$$

We then compute the traces over E using the specific form of V in Eq. (5.73): For instance,

$$\begin{aligned}\mathrm{tr}_E(V^2\rho_S\rho_E) &= \mathrm{tr}_E(V^2\rho_E)\rho_S\\ &= g^2\,\mathrm{tr}_E\Big\{(L^\dagger L\sigma_-^E\sigma_+^E + LL^\dagger\sigma_+^E\sigma_-^E)\rho_E\Big\}\rho_S\\ &= g^2\Big\{\langle\sigma_-^E\sigma_+^E\rangle L^\dagger L\rho_S + \langle\sigma_+^E\sigma_-^E\rangle LL^\dagger\rho_S\Big\}.\end{aligned}$$

Using Eq. (5.72) we get

$$\langle\sigma_-^E\sigma_+^E\rangle = 1-f, \qquad \langle\sigma_+^E\sigma_-^E\rangle = f. \tag{5.77}$$

Thus,

$$\mathrm{tr}_E(V^2\rho_S\rho_E) = g^2\Big\{(1-f)L^\dagger L\rho_S + fLL^\dagger\rho_S\Big\}.$$

By symmetry, we expect that

$$\mathrm{tr}_E(\rho_S\rho_E V^2) = g^2\Big\{(1-f)\rho_S L^\dagger L + f\rho_S LL^\dagger\Big\}.$$

Finally, we compute

$$\begin{aligned}\mathrm{tr}_E(V\rho_S\rho_E V) &= g^2\,\mathrm{tr}_E\Big\{(L^\dagger\sigma_-^E + L\sigma_+^E)\rho_S\rho_E(L^\dagger\sigma_-^E + L\sigma_+^E)\Big\}\\ &= g^2\Big\{L^\dagger\rho_S L\,\mathrm{tr}_E(\sigma_-^E\rho_E\sigma_+^E) + L\rho_S L^\dagger\,\mathrm{tr}_E(\sigma_+^E\rho_E\sigma_-^E) + L^\dagger\rho_S L^\dagger\,\mathrm{tr}_E(\sigma_-^E\rho_E\sigma_-^E) + L\rho_S L\,\mathrm{tr}_E(\sigma_+^E\rho_E\sigma_+^E)\Big\}.\end{aligned}$$

The last two terms are zero because $(\sigma_\pm^E)^2 = 0$. In the first two terms, we can now use the cyclic property of the trace, which yields

$$\mathrm{tr}_E(V\rho_S\rho_E V) = g^2\Big\{fL^\dagger\rho_S L + (1-f)L\rho_S L^\dagger\Big\}.$$

Combining everything in Eq. (5.76) then finally leads to

$$\boxed{D(\rho_S) = g^2\tau(1-f)\Big[L\rho_S L^\dagger - \frac{1}{2}\{L^\dagger L,\rho_S\}\Big] + g^2\tau\, f\Big[L^\dagger\rho_S L - \frac{1}{2}\{LL^\dagger,\rho_S\}\Big].} \tag{5.78}$$

We therefore recognize here a Linbdlad master equation with jump operators L and $L^\dagger$ and rates $\gamma_- = g^2\tau(1-f)$ and $\gamma_+ = g^2\tau f$.

Collisional models are neat because they give you substantial control over the types of master equations you can construct. Eq. (5.78) provides a very general recipe. Given any system S, if you want to couple it to an environment with an interaction of the form (5.73), this will produce a dissipator $D(\rho_S)$ having both L and $L^\dagger$ as jump operators.

Chapter 6

Continuous variables

Continuous variables is a fancy name we give to harmonic oscillators. So far we have talked about systems with a finite number of states, so that everything is discrete. Now we will talk about harmonic oscillators, which have an infinite number of levels. Of course, these levels are also discrete. However, it turns out that many things can be described in terms of continuous variables, such as coherent states and the quantum phase space representation.

Continuous variables systems occur naturally in many platforms. The most important example is quantum optics, where, it turns out, the quantum properties of the electromagnetic field can be represented in terms of harmonic oscillators. Other continuous variable platforms include trapped ions, nano- or micro-mechanical oscillators and Bose-Einstein condensates.

This chapter provides a first look into continuous variables. After this, we will start to work with both discrete and continuous variable systems, side by side. More advanced properties will be discussed later, or can be found in the excellent book by Alessio Serafini entitled *"Quantum Continuous Variables"*.

6.1 Creation and annihilation operators

The starting point of continuous variable systems is an operator a called the *annihilation operator* and its Hermitian conjugate $a^\dagger$, called the *creation operator*. They are defined so as to satisfy the following algebra:

$$\boxed{[a, a^\dagger] = 1.} \tag{6.1}$$

All properties of these operators and the Hilbert space they represent follow from this simple commutation relation, as we will see below. Another set of operators which can be used as the starting point of the discussion are the *position and momentum operators*

q and p. They satisfy

$$\boxed{[q,p]=i.} \tag{6.2}$$

In quantum optics, they no longer represent position and momentum, but are related to the electric and magnetic fields. In this case they are usually called *quadrature operators*. We define q and p to be dimensionless. Then they are related to the creation and annihilation operators according to

$$\begin{aligned} q &= \frac{1}{\sqrt{2}}(a^\dagger + a) & & & a &= \frac{1}{\sqrt{2}}(q+ip) \\ & & \iff & & & \\ p &= \frac{i}{\sqrt{2}}(a^\dagger - a) & & & a^\dagger &= \frac{1}{\sqrt{2}}(q-ip)\,. \end{aligned} \tag{6.3}$$

From this it can be clearly seen that q and p are Hermitian operators, even though a is not. Also, please take a second to verify that with this relation Eq. (6.1) indeed implies (6.2) and vice-versa.

Mechanical oscillators

The operators a, $a^\dagger$, q and p appear in two main contexts: mechanical oscillators and second quantization. The latter will be discussed below. A mechanical oscillator is specified by the Hamiltonian

$$H = \frac{P^2}{2m} + \frac{1}{2}m\omega^2 Q^2, \tag{6.4}$$

where m is the mass and ω is the frequency. Moreover Q and P are the position and momentum operators satisfying

$$[Q,P] = i\hbar. \tag{6.5}$$

Now define the dimensionless operators

$$q = \sqrt{\frac{m\omega}{\hbar}}Q, \qquad p = \frac{P}{\sqrt{m\hbar\omega}}. \tag{6.6}$$

Then Eq. (6.5) implies that q and p will satisfy (6.2). In terms of q and p, the Hamiltonian (6.4) becomes

$$H = \frac{\hbar\omega}{2}(p^2+q^2), \tag{6.7}$$

which, you have to admit, is way more elegant than (6.4). Using now Eq. (6.3) we finally write the Hamiltonian as

$$H = \hbar\omega(a^\dagger a + 1/2). \tag{6.8}$$

Eqs. (6.7) and (6.8) show very well why $\hbar$ is not important: it simply redefines the energy scale. If we set $\hbar = 1$, as we shall henceforth do, we are simply measuring energy in units of frequency.

In the days of Schrödinger, harmonic oscillators were usually used either as toy models or as an effective description of some other phenomena such as, for instance, the vibration of molecules. In the last two decades this has change and we are now able to observe quantum effects in actual mechanical mesoscopic- (nano- or micro-) oscillators. This is usually done by engineering thin suspended membranes, which can then undergo mechanical vibrations. This field is usually known as **optomechanics** since most investigations involve the contact of the membranes with radiation. I find it absolutely fascinating that in our day and age we can observe quantum effects as awesome as entanglement and coherence in these mechanical objects. I love the century we live in!

An algebraic problem

In Eq. (6.8) we see the appearance of the Hermitian operator $a^\dagger a$, called the **number operator**. To find the eigenstuff of H we therefore only need to know the eigenstuff of $a^\dagger a$. We have therefore arrived at a very clean mathematical problem: *given a non-Hermitian operator a, satisfying $[a, a^\dagger] = 1$, find the eigenvalues and eigenvectors of $a^\dagger a$.* This is a really important problem that appears often in all areas of quantum physics: given an algebra, find the eigenstuff. Maybe you have seen this before, but I will nonetheless do it again, because I think this is one of those things that everyone should know.

Here we go. Since $a^\dagger a$ is Hermitian, its eigenvalues must be real and its eigenvectors can be chosen to form an orthonormal basis. Let us write them as

$$\boxed{a^\dagger a|n\rangle = n|n\rangle.} \tag{6.9}$$

Our goal is to find the allowed n and the corresponding $|n\rangle$. The first thing we notice is that $a^\dagger a$ must be positive semi-definite operator, so n cannot be negative:

$$n = \langle n|a^\dagger a|n\rangle \geq 0.$$

Next we use Eq. (6.1) to show that

$$\boxed{[a^\dagger a, a] = -a, \qquad\qquad [a^\dagger a, a^\dagger] = a^\dagger.} \tag{6.10}$$

This type of structure is a signature of a *ladder* like spectrum (that is, when the eigenvalues are equally spaced). To see that, we use these commutation relations to compute:

$$(a^\dagger a)a|n\rangle = [a(a^\dagger a) - a]|n\rangle = a(a^\dagger a - 1)|n\rangle = (n-1)a|n\rangle.$$

Hence, we conclude that if $|n\rangle$ is an eigenvector with eigenvalue n, then $a|n\rangle$ is also an eigenvector, but with eigenvalue $(n-1)$ [This is the key argument. Make sure you understand what this sentence means.]. However, I wouldn't call this $|n-1\rangle$ just yet because $a|n\rangle$ is not normalized. Thus we need to write

$$|n-1\rangle = \gamma a|n\rangle,$$

where γ is a normalization constant. To find it we simply write

$$1 = \langle n-1|n-1\rangle = |\gamma|^2 \langle n|a^\dagger a|n\rangle = |\gamma|^2 n.$$

Thus $|\gamma|^2| = 1/n$. The actual sign of γ is arbitrary so we choose it for simplicity as being real and positive. We then get

$$|n-1\rangle = \frac{a}{\sqrt{n}}|n\rangle.$$

From this analysis we conclude that a reduces the eigenvalues by unity:

$$a|n\rangle = \sqrt{n}|n-1\rangle.$$

We can do a similar analysis with $a^\dagger$. We again use Eq. (6.10) to compute

$$(a^\dagger a)a^\dagger|n\rangle = (n+1)a^\dagger|n\rangle.$$

Thus $a^\dagger$ raises the eigenvalue by unity. The normalization factor is found by a similar procedure: we write $|n+1\rangle = \beta a^\dagger|n\rangle$, for some constant β, and then compute

$$1 = \langle n+1|n+1\rangle = |\beta|^2\langle n|aa^\dagger|n\rangle = |\beta|^2\langle n|(1+a^\dagger a)|n\rangle = |\beta|^2(n+1).$$

Thus

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle.$$

These results are important, so let me summarize them in a boxed equation:

$$\boxed{a|n\rangle = \sqrt{n}|n-1\rangle, \qquad a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle}. \tag{6.11}$$

From this formula we can see why the operators a and $a^\dagger$ also receive the name *lowering and raising operators*.

Now comes the trickiest (and most beautiful) argument. We have seen that if n is an eigenvalue, then $n \pm 1$, $n \pm 2$, etc., will all be eigenvalues. But this doesn't mean that n itself should be an integer. Maybe we find one eigenvalue which is 42.42 so that the eigenvalues are 41.42, 43.42 and so on. Of course, you know that is not true and n must be integer. To show that, we proceed as follows. Suppose we start with some eigenstate $|n\rangle$ and keep on applying a a bunch of times. At each application we will lower the eigenvalue by one tick:

$$a^\ell|n\rangle = \sqrt{n(n-1)\ldots(n-\ell+1)}|n-\ell\rangle.$$

But this crazy party cannot continue forever because, as we have just discussed, the eigenvalues of $a^\dagger a$ cannot be negative. They can, at most, be zero. The only way for this to happen is if there exists a certain integer ℓ for which $a^\ell|n\rangle \neq 0$ but $a^{\ell+1}|n\rangle = 0$. And this can only happen if $\ell = n$ because, then

$$a^{\ell+1}|n\rangle = \sqrt{n(n-1)\ldots(n-\ell+1)(n-\ell)}|n-\ell-1\rangle = 0,$$

and the term $n - \ell$ will vanish. Since ℓ is an integer, we therefore conclude that n must also be an integer. Thus, we finally conclude that

$$\text{eigs}(a^\dagger a) = n \in \{0, 1, 2, 3, \ldots\}. \tag{6.12}$$

It is for this reason that $a^\dagger a$ is called the number operator: we usually say $a^\dagger a$ counts the number of quanta in a given state: given a state $|n\rangle$, you first apply a to annihilate one quant and then $a^\dagger$ to create it back again. The proportionality factor is the eigenvalue n. Curiously, this analysis seem to imply that if you want to count how many people there are in a room, you first need to annihilate one person and then create a fresh new human. Quantum mechanics is indeed strange.

This analysis also serves to define the state with $n = 0$, which we call the **vacuum**, $|0\rangle$. It is defined by

$$\boxed{a|0\rangle = 0.} \tag{6.13}$$

We can build all states starting from the vacuum and applying $a^\dagger$ successively:

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle. \tag{6.14}$$

Using this and the algebra of a and $a^\dagger$ it then follows that the states $|n\rangle$ form an orthonormal basis, as expected:

$$\langle n|m\rangle = \delta_{n,m}.$$

The states $|n\rangle$ are called **Fock states**, although this nomenclature is more correctly employed in the case of multiple modes, as we will now discuss.

Multiple modes and second quantization

It is straightforward to generalize the idea of creation and annihilation operators to composite systems. We simply define a set of annihilation operators a_i, where $i = 1, 2, \ldots, N$. It is customary to use the word **mode** to label each i. Thus we say things like "mode a_2". These operators are defined to satisfy

$$[a_i, a_j^\dagger] = \delta_{i,j}, \qquad [a_i, a_j] = 0. \tag{6.15}$$

That is, a_i with $a_i^\dagger$ behaves just like before, whereas a_i with $a_j^\dagger$ commute if $j \neq i$. Moreover annihilation operators always commute among themselves. Taking the adjoint of $[a_i, a_j] = 0$ we see that the same will be true for the creation operators $[a_i^\dagger, a_j^\dagger] = 0$. Using the same transformation as in Eq. (6.3), but with indices everywhere, we can also define quadrature operators q_i and p_i, which will then satisfy

$$[q_i, p_j] = i\delta_{i,j}, \qquad [q_i, q_j] = [p_i, p_j] = 0. \tag{6.16}$$

Multi-mode systems can appear in mechanical contexts. For instance, consider two mechanical oscillators coupled by springs, as in Fig. 6.1. Each oscillator has a natural

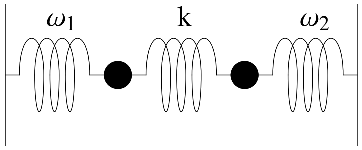

Figure 6.1: Two harmonic oscillators coupled by a harmonic spring.

frequency ω_1 and ω_2 and they are coupled by a spring constant k. Assuming unit mass, the Hamiltonian will then be

$$H = \left(p_1^2 + \frac{\omega_1^2}{2} q_1^2\right) + \left(p_2^2 + \frac{\omega_2^2}{2} q_2^2\right) + \frac{k}{2}(q_1 - q_2)^2. \tag{6.17}$$

If we want we can also transform this into a_i and $a_i^\dagger$, or we can extend it to multiple oscillators forming a chain. In fact, these "harmonic chains" are a widely studied topic in the literature because they can always be solved analytically and they are the starting point for a series of interesting quantum effects. We will have the opportunity to practice with some of these solutions later on.

But by far the most important use of multi-mode systems is in **second quantization**. Since operators pertaining to different modes commute, the Hilbert space of a multi-mode system will be described by a basis

$$|\boldsymbol{n}\rangle = |n_1, n_2, \dots, n_N\rangle, \qquad n_i = 0, 1, 2, \dots. \tag{6.18}$$

These are called **Fock states** and are the eigenstates of the number operators $a_i^\dagger a_i$:

$$a_i^\dagger a_i |\boldsymbol{n}\rangle = n_i |\boldsymbol{n}\rangle. \tag{6.19}$$

Thus, $a_i^\dagger a_i$ counts the number of quanta in mode i.

Second quantization is essentially a change of perspective from "quanta" to "particles". After all, what the hell is a quanta anyway? In second quantization we say $a_i^\dagger a_i$ is the operator counting the number of *particles* in mode i. Then $a_i^\dagger$ is the operator which creates a particle in mode i, whereas $a_i^\dagger$ annihilates. You may also be wondering what is a "mode" in this case. Well, there is actually an infinite number of choices. We could take for instance $i = x$, the position in space. Then $a_x^\dagger$ is the operator which creates a particle at position x. In quantum field theory we call it $\psi^\dagger(x)$ instead. But it's the same thing.

According to Eq. (6.19) each mode can have an arbitrary number n of particles. We then call a_i a **bosonic mode**. So whenever someone says "consider a set of bosonic modes" they mean a set of operators a_i satisfying (6.15). This is to be contrasted with Fermionic systems, for which the only allowed Fock states are $n = 0$ and $n = 1$ (due to the Pauli exclusion principle). We will not discuss much of fermionic systems in this course, but the idea is somewhat similar. We also define creation and annihilation operators, except that now they satisfy a different algebra:

$$\{c_i, c_j^\dagger\} = \delta_{i,j}, \qquad \{c_i, c_j\} = 0, \tag{6.20}$$

where $\{A, B\} = AB + BA$ is the anti-commutator. If we repeat the diagonalization procedure of the last section for this kind of algebra we will find a similar "Fock structure" but with the only allowed eigenvalues being $n_i = 0$ and $n_i = 1$.

The most important bosonic system is the electromagnetic field. The excitations are then the photons and the modes are usually chosen to be the momentum and polarization. Hence, we usually write an annihilation operator as $a_{\boldsymbol{k},\lambda}$ where $\boldsymbol{k} = (k_x, k_y, k_z)$ is the momentum and $\lambda = \pm 1$ is the polarization. Moreover, the Hamiltonian of the electromagnetic field is written as

$$H = \sum_{\boldsymbol{k},\lambda} \omega_{\boldsymbol{k}} a^\dagger_{\boldsymbol{k},\lambda} a_{\boldsymbol{k},\lambda}, \tag{6.21}$$

where $\omega_{\boldsymbol{k}}$ is the frequency of each mode and is given by[1] $\omega_{\boldsymbol{k}} = c|\boldsymbol{k}|$ where c is the speed of light.

You have noticed that my discussion of second quantization was rather shallow. I apologize for that. But I have to do it like this, otherwise we would stray too far. Second quantization is covered in many books on condensed matter, quantum many-body and quantum field theory. A book which I really like is Feynman's *Statistical mechanics: a set of lectures*".

6.2 Some important Hamiltonians

In this section we briefly discuss some important Hamiltonians that appear often in controlled quantum experiments.

Optical cavities

Many controlled experiments take place inside optical cavities, like the one represented in my amazing drawing in Fig. 6.2 (it took me 30 minutes to draw it!). The cavity is made up of highly reflective mirrors allowing the photons to survive for some time, forming standing wave patterns. Unlike in free space, where all radiation modes can exist equally, the confinement inside the cavity favors those radiation modes whose frequencies are close to the *cavity frequency* ω_c, which is related to the geometry of the cavity. It is therefore common to consider only one radiation mode, with operator a and frequency ω_c.

The photons always have a finite lifetime so more photons need to be injected all the time. This is usually done by making one of the mirrors semi-transparent and pumping it with a laser from the outside, with frequency ω_p. Of course, since photons can come in, they can also leak out. This leakage is an intrinsically irreversible process and can only be described using the theory of open quantum systems, which we will get to in the next chapter. Hence, we will omit the process of photon losses for now. The Hamiltonian describing a single mode pumped externally by a laser then has the form

$$H = \omega_c a^\dagger a + \epsilon a^\dagger e^{-i\omega_p t} + \epsilon^* a e^{i\omega_p t}, \tag{6.22}$$

---

[1] If we define $\omega = 2\pi\nu$ and $|\boldsymbol{k}| = 2\pi/\lambda$ we see that this is nothing but the relation $\nu = \lambda c$ that you learned in high school.

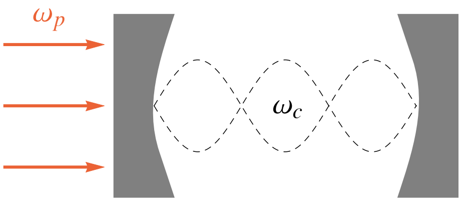

Figure 6.2: An optical cavity of frequency ω_c, pumped from the outside by a laser of frequency ω_p.

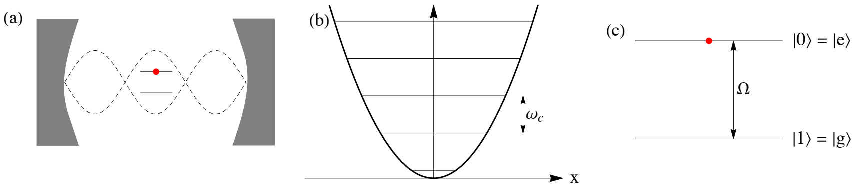

Figure 6.3: (a) Typical scenario for light-matter interaction: an atom, modeled as a two-level system, is placed inside a cavity in which there is only one cavity mode. The atom then absorbs and emits photons jumping up and down from the ground-state to the excited state. (b) The cavity field is represented by a harmonic oscillator of frequency ω_c. (c) The atom is represented as a two-level system (qubit) with energy gap Ω. When the atom Hamiltonian is $+\sigma_z$ then the ground-state will be $|1\rangle$ and the excited state will be $|0\rangle$.

where ϵ is the pump amplitude and is related to the laser power P according to $|\epsilon|^2 = \gamma P/\hbar\omega_p$ where γ is the cavity loss rate (the rate at which photons can go through the semi-transparent mirror). This Hamiltonian is very simple, but is time-dependent. Lucky for us, however, this time dependence can be eliminated using the concept of a rotating frame, as will be discussed below.

Jaynes-Cummings and Rabi models

Quantum information has always been intimately related with quantum optics and atomic physics, so *light-matter interaction* is an essential topic in the field. The two most important models in this sense are the Jaynes-Cummings and Rabi models, both of which describe the interaction of a single radiation mode with a single atom, approximated as a two-level system. The basic idea of both models is the exchange of quanta between the two systems; that is, sometimes the atom absorbs a photon and jumps to an excited state and sometimes it emits a photon and drops down to the ground-state. These effects of course take place on free space, but we are usually interested in controlled experiments performed inside optical cavities. The situation is then like that of

Fig. 6.3.

The Jaynes-Cummings model reads

$$H = \omega a^\dagger a + \frac{\Omega}{2}\sigma_z + \lambda(a\sigma_+ + a^\dagger\sigma_-). \tag{6.23}$$

The first two terms are the free Hamiltonians of the cavity field, with frequency ω_c, and the atom, with energy gap Ω. Whenever the atom Hamiltonian is written as $+\sigma_z$, the ground-state will be $|g\rangle = |1\rangle$ and the excited state will be $|e\rangle = |0\rangle$ [see Fig. 6.3(c)]. Finally, the last term in (6.23) is the light-atom coupling. The term $a\sigma_+$ describes the process where a photon is annihilated and the atom jumps to the excited state. Similarly, $a^\dagger\sigma_-$ describes the opposite process. The Hamiltonian must always be Hermitian so every time we include a certain type of process, we must also include its reverse.

The type of interaction in Eq. (6.23) introduces a special *symmetry* to the problem. Namely, it conserves the number of quanta in the system:

$$[H, a^\dagger a + \sigma_z] = 0. \tag{6.24}$$

This means that if you start the evolution with 7 photons and the atom in the ground-state, then at all times you will either have 7 photons + ground-state or 6 photons and the atom in the excited state. This is a very special symmetry and is the reason why the Jaynes-Cummings model turns out to be easy to deal with.

However, if we start with a physical derivation of the light-atom interaction, we will see that it is not exactly like the Jaynes-Cummings Hamiltonian (6.23). Instead, it looks more like the Rabi model

$$H = \omega a^\dagger a + \frac{\Omega}{2}\sigma_z + \lambda(a + a^\dagger)\sigma_x. \tag{6.25}$$

The difference is only in the last term. In fact, if we recall that $\sigma_x = \sigma_+ + \sigma_-$, we get

$$(a + a^\dagger)\sigma_x = (a\sigma_+ + a^\dagger\sigma_-) + (a^\dagger\sigma_+ + a\sigma_-).$$

The first term in parenthesis is exactly the Jaynes-Cummings interaction, so the new thing here is the term $(a^\dagger\sigma_+ + a\sigma_-)$. It describes a process where the atom jumps to the excited state *and* emits a photon, something which seems rather strange at first. Moreover, this new term destroys the pretty symmetry (6.24), making the Rabi model much more complicated to deal with, but also much richer from a physical point of view. Notwithstanding, as we will see below, if λ is small compared to ω_c, Ω this new term becomes negligible and the Rabi model approximately tends to the JC Hamiltonian.

6.3 Rotating frames and interaction picture

In this section I want to introduce the concept of *rotating frames*, which is a small generalization of the interaction and Heisenberg pictures that you may have learned in quantum mechanics. Consider a system with density matrix ρ evolving according to von Neumann's equation (we could do the same with Schrödinger's equation)

$$\frac{d\rho}{dt} = -i[H(t), \rho], \tag{6.26}$$

where $H(t)$ is a possibly time-dependent Hamiltonian. We can always move to a *rotating frame* by defning a new density matrix

$$\tilde{\rho}_t = S(t)\rho S^\dagger(t), \tag{6.27}$$

where $S(t)$ is an arbitrary unitary. I will leave to you as an exercise to show that $\tilde{\rho}$ will also obey a von Neumann equation

$$\frac{\mathrm{d}\tilde{\rho}}{\mathrm{d}t} = -i[\tilde{H}(t), \tilde{\rho}], \tag{6.28}$$

but with an effective Hamiltonian[2]

$$\boxed{\tilde{H}(t) = i\frac{\mathrm{d}S}{\mathrm{d}t}S^\dagger + SHS^\dagger.} \tag{6.29}$$

Thus, we see that in any rotating frame the system always obeys von Neumann's (or Schrödinger's) equation, but the Hamiltonian changes from $H(t)$ to $\tilde{H}(t)$. Note that this result is absolutely general and holds for any unitary $S(t)$. Of course, whether it is useful or not will depend on your smart choice for $S(t)$.

Before we move to applications, I need to mention that computing the first term in Eq (6.29) can be tricky. Usually we write unitaries as $S(t) = e^{iK(t)}$ where K is Hermitian. Then, one may easily verify the following BCH expansion

$$\boxed{\frac{\mathrm{d}e^{iK}}{\mathrm{d}t}e^{-iK} = i\frac{\mathrm{d}K}{\mathrm{d}t} + \frac{i^2}{2}[K, \frac{\mathrm{d}K}{\mathrm{d}t}] + \frac{i^3}{3!}[K,[K,\frac{\mathrm{d}K}{\mathrm{d}t}]] + \dots.} \tag{6.30}$$

The important point here is whether or not K commutes with $\mathrm{d}K/\mathrm{d}t$. If that is the case then only the first term survives and things are easy and pretty. Otherwise, you may get an infinite series. I strongly recommend you always use this formula, because then you are always sure you will not get into trouble.

Eliminating time-dependences

A simple yet useful application of rotating frames is to eliminate the time-dependence of certain simple Hamiltonians, such as the pumped cavity (6.22). In this case the unitary that does the job is

$$S(t) = e^{i\omega_p t a^\dagger a}. \tag{6.31}$$

---

[2] To derive this equation it is necessary to use the following trick: since $SS^\dagger = 1$ then

$$0 = \frac{\mathrm{d}SS^\dagger}{\mathrm{d}t} = S\frac{\mathrm{d}S^\dagger}{\mathrm{d}t} + \frac{\mathrm{d}S}{\mathrm{d}t}S^\dagger \qquad \longrightarrow \qquad S\frac{\mathrm{d}S^\dagger}{\mathrm{d}t} = -\frac{\mathrm{d}S}{\mathrm{d}t}S^\dagger.$$

That is, we move to a frame that is rotating at the same frequency as the pump laser ω_p. Using the BCH expansion (1.70) one may show that

$$\boxed{e^{i\alpha a^\dagger a} a e^{-i\alpha a^\dagger a} = e^{-i\alpha} a, \qquad e^{i\alpha a^\dagger a} a^\dagger e^{-i\alpha a^\dagger a} = e^{i\alpha} a^\dagger,} \tag{6.32}$$

which are easy to remember: a goes with negative α and $a^\dagger$ with positive α. It then follows that

$$S(t)\Big[\epsilon a^\dagger e^{-i\omega_p t} + \epsilon a e^{i\omega_p t}\Big] S^\dagger(t) = \epsilon a^\dagger + \epsilon a,$$

while $S(t)$ has no effect on $a^\dagger a$. Moreover, this is one of those cases where only the first term in (6.30) contributes:

$$\frac{\mathrm{d}S}{\mathrm{d}t} S^\dagger = i\omega_p a^\dagger a.$$

Thus Eq. (6.29) becomes

$$\boxed{\tilde{H} = (\omega_c - \omega_p) a^\dagger a + \epsilon a^\dagger + \epsilon^* a.} \tag{6.33}$$

We therefore conclude that in this rotating frame the Hamiltonian is time-independent, but evolves according to the **detuned frequency** $\Delta = \omega_c - \omega_p$. This idea of *detuning* a frequency is extremely important in quantum optics applications since it is an easy way to change the parameters in the problem.

For more general bosonic Hamiltonians containing a pump term, the time-dependence can be eliminated by the same transformation, provided the remainder of the Hamiltonian conserves the number of quanta (i.e., when all operators have an equal number of as and $a^\dagger$s). This is due to the simple rule imposed by (6.32), which says that every a gets a term $e^{-i\omega_p t}$ while every $a^\dagger$ gets a $e^{i\omega_p t}$. Thus, a Hamiltonian such as

$$H = \omega_a^\dagger a + \frac{U}{2} a^\dagger a^\dagger a a + \epsilon a^\dagger e^{-i\omega_p t} + \epsilon^* a e^{i\omega_p t},$$

will lead to a rotating frame Hamiltonian

$$\tilde{H} = (\omega - \omega_p) a^\dagger a + \frac{U}{2} a^\dagger a^\dagger a a + \epsilon a^\dagger + \epsilon^* a.$$

Once you get the hang of it, it is quite easy: detune the frequency and get rid of the exponential. But be careful. This can only be done if the number of as and $a^\dagger$s is the same. For instance,

$$H = \omega_a^\dagger a + \chi (a + a^\dagger)^4 + \epsilon a^\dagger e^{-i\omega_p t} + \epsilon^* a e^{i\omega_p t},$$

would not have a time-independent rotating frame under the transformation (6.31) because if you expand $(a + a^\dagger)^4$ there will be terms with a unbalanced number of as and $a^\dagger$s.

A similar rotating frame transformation also works for qubit systems of the form

$$H = \frac{\Omega}{2}\sigma_z + \frac{\lambda}{2}(\sigma_+ e^{-i\omega_p t} + \sigma_- e^{i\omega_p t}) \tag{6.34}$$

$$= \frac{\Omega}{2}\sigma_z + \frac{\lambda}{2}[\sigma_x \cos(\omega_p t) + \sigma_y \sin(\omega_p t)]. \tag{6.35}$$

This Hamiltonian appears often in magnetic resonance because it represents a spin 1/2 particle subject to a constant field Ω in the z direction and a rotating field λ in the xy plane. Remarkably, the transformation here is almost exactly as in the bosonic case:

$$S(t) = e^{i\omega_p t \sigma_z/2}. \tag{6.36}$$

In this case the idea of a rotating frame becomes a bit more intuitive: the Hamiltonian is time-dependent because there is a field rotating in the xy plane. So to get rid of it, we go to a frame that is rotating around the z axis by an angle $\omega_p t$. I will leave for you to check that this $S(t)$ indeed does the job. One thing that is useful to know is that Eq. (6.32) is translated almost literally to the spin case:

$$e^{i\alpha\sigma_z/2}\sigma_- e^{-i\alpha\sigma_z/2} = e^{-i\alpha}\sigma_-, \qquad e^{i\alpha\sigma_z/2}\sigma_+ e^{-i\alpha\sigma_z/2} = e^{i\alpha}\sigma_+, \tag{6.37}$$

Interaction picture

Now let us consider another scenario. Suppose the Hamiltonian is time-independent but can be written in the standard perturbation-theory-style

$$H = H_0 + V, \tag{6.38}$$

where H_0 is easy to handle but V is not. Then choose

$$S(t) = e^{iH_0 t}. \tag{6.39}$$

Eq. (6.29) then becomes

$$\boxed{\tilde{H}(t) = S(t) V S^\dagger(t).} \tag{6.40}$$

This is the interaction picture: we eliminate the dependence on H_0 at the cost of transforming a time-independent Hamiltonian $H_0 + V$ into a time-dependent Hamiltonian $SVS^\dagger$.

The interaction picture is usually employed as the starting point of time-dependent perturbation theory. We will learn a bit more about this below. But to get a first glimpse, consider the Rabi Hamiltonian (6.25) and let us move to the interaction picture with respect to $H_0 = \omega a^\dagger a + \frac{\Omega}{2}\sigma_z$. Using Eqs. (6.32) and (6.37) we then find

$$\tilde{H}(t) = \lambda\Big[a\sigma_+ e^{i(\Omega-\omega_c)t} + a^\dagger\sigma_- e^{-i(\Omega-\omega_c)t}\Big] + \lambda\Big[a^\dagger\sigma_+ e^{i(\Omega+\omega_c)t} + a\sigma_- e^{-i(\Omega+\omega_c)t}\Big]. \tag{6.41}$$

In the interaction picture we see more clearly the difference between the two types of couplings. The first term, which is the Jaynes-Cummings coupling, oscillates in time with a frequency $\Omega - \omega_c$, which will be very small when Ω is close to ω_c. The second term, on the other hand, oscillates quickly with frequency $\omega_c + \Omega$, which is in general a much faster frequency that $\omega_c - \Omega$. We therefore see the appearance of two time scales, the JC term, which is slow, and the Rabi dude which give rise to fast oscillations.

Eq. (6.41) is frequently used as the starting point to justify why sometimes we can throw away the last term (and hence obtain the Jaynes-Cummings model (6.23) from the Rabi model). The idea is called the **rotating-wave approximation (RWA)** and is motivated by the fact that if $\Omega + \omega$ is very large, the last terms will oscillate rapidly around zero average and hence will have a small contribution to the dynamics. But this explanation is only partially convincing, so be careful. In the end of the day, the RWA is really an argument on time-dependent perturbation theory. Hence, it will only be good when λ is small compared to ω_c and Ω. Thus, the RWA is better stated as follows: *if* $\lambda \ll \omega_c, \Omega$ *and* $\omega_c \sim \Omega$*, it is reasonable to throw away the fast oscillating terms in the interaction picture*. For an interesting discussion connection with perturbation theory, see the Appendix in `arXiv 1601.07528`.

Heisenberg picture

In the interaction picture we started with a Hamiltonian $H = H_0 + V$ and went to a rotating frame with H_0. In the Heisenberg picture, we go all the way through. That is, we go to a rotating frame (6.29) with $S(t) = e^{iHt}$. For now I will assume H is time-independent, but the final result also holds in the time-dependent case. As a result we find

$$\tilde{H} = 0 \tag{6.42}$$

Consequently, the solution of the rotating frame Eq. (6.28) will be simply

$$\tilde{\rho}(t) = \tilde{\rho}(0) = \rho(0). \tag{6.43}$$

But by Eq. (6.27) we have $\tilde{\rho}(t) = S(t)\rho(t)S^\dagger(t)$ so we get

$$\rho(t) = S^\dagger(t)\rho(0)S(t) = e^{-iHt}\rho(0)e^{iHt}. \tag{6.44}$$

You may now be thinking “DUH! This is is just the solution of the of von Neumann’s equation!”. Yes, that’s exactly the point. The solution of von Neumann’s equation is exactly that special rotating frame where time stands still (like in the Rush song!).

In the Heisenberg picture we usually transfer the time-dependence to the operators, instead of the states. Recall that given an arbitrary operator A, its expectation value will be $\langle A\rangle = \mathrm{tr}(A\rho)$. Using Eq. (6.44) we then get

$$\langle A\rangle = \mathrm{tr}\left\{Ae^{-iHt}\rho(0)e^{iHt}\right\} = \mathrm{tr}\left\{e^{iHt}Ae^{iHt}\rho(0)\right\}. \tag{6.45}$$

This formula summarizes well the Schrödinger vs. Heisenberg ambiguity. It provides two equivalent ways to compute $\langle A\rangle$. In the first, which is the usual Schrödinger picture approach, the state $\rho(t)$ evolves in time and A is time-independent. In the second, the state ρ is fixed at $\rho(0)$ and we transfer the time evolution to the operator. It is customary to define the *Heisenberg operator*

$$A_H(t) = A(t) = e^{iHt}Ae^{-iHt}. \tag{6.46}$$

Some people write $A_H(t)$ to emphasize that this is different from A What I usually do is just be careful to always write the time argument in $A(t)$.

By direct differentiation one may verify that the operator $A(t)$ satisfies the **Heisenberg equation**

$$\boxed{\frac{\mathrm{d}A(t)}{\mathrm{d}t} = i[H, A(t)].} \tag{6.47}$$

This is to be interpreted as an equation for the evolution of the operator $A(t)$. If what you are interested is instead the evolution of the expectation value $\langle A\rangle_t$, then it doesn't matter which picture you use. In the Heisenberg picture, Eq. (6.47) directly gives you

$$\boxed{\frac{\mathrm{d}\langle A\rangle}{\mathrm{d}t} = i\langle[H, A]\rangle.} \tag{6.48}$$

But you can also get the same equation in the Schrödinger picture using the von Neumann equation:

$$\frac{\mathrm{d}\langle A\rangle}{\mathrm{d}t} = \mathrm{tr}\left\{A\frac{\mathrm{d}\rho}{\mathrm{d}t}\right\} = -i\,\mathrm{tr}\left\{A[H,\rho]\right\} = i\,\mathrm{tr}\left\{[H,A]\rho\right\},$$

where, in the last line, all I did was rearrange the commutator using the cyclic property of the trace.

About time-dependent Hamiltonians

The solution of Schrödinger's or von Neumann's equation for time-independent Hamiltonians is very easy, being simply e^{-iHt}. However, when the Hamiltonian is time-dependent this solution no longer works. Let us then see how to write down the solution in this case. I will do so for the case of Schödinger's equation, simply because it looks a little bit cuter. It is straightforward to generalized to von Neumann's equation. Our starting point is thus the equation

$$\partial_t|\psi_t\rangle = -iH(t)|\psi_t\rangle. \tag{6.49}$$

In order to figure out what the solution will be in this case, we follow the maxim of Polish mathematician Marc Kac: **"be wise, discretize!"** That is, we assume that the Hamiltonian $H(t)$ is actually piecewise constant at intervals Δt, having the value $H(n\Delta t)$ during the interval between $n\Delta t$ and $(n+1)\Delta t$ (something like what is shown in Fig. 6.4, but for the operator $H(t)$). We can then solve Eq. (6.49) exactly for one interval:

$$|\psi((n+1)\Delta t)\rangle = e^{-i\Delta t H(n\Delta t)}|\psi(n\Delta t)\rangle. \tag{6.50}$$

From this we can proceed sequentially, using the solution for a given interval as the initial condition for the next. This allows us to glue together a solution between $t_0 = M\Delta t$ and $t = (N+1)\Delta t$ (with M, N integers and $N > M$):

$$|\psi_t\rangle = \left[e^{-i\Delta t H(N\Delta t)}e^{-i\Delta t H((N-1)\Delta t)}\dots e^{-i\Delta t H(M\Delta t)}\right]|\psi_{t_0}\rangle. \tag{6.51}$$

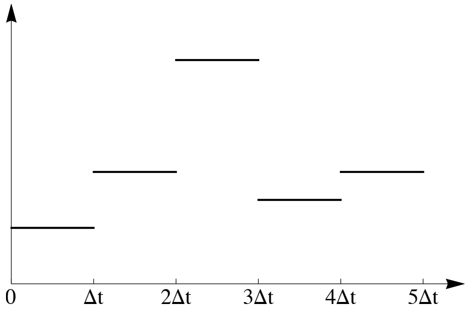

Figure 6.4: A silly example of a piecewise constant function.

Of course, this discretization is just a trick. We can now take $\Delta t \to 0$ and we will have solved for the most general time-dependent Hamiltonian.

If we define the time-evolution operator according to

$$|\psi_t\rangle = U(t, t_0)|\psi_{t_0}\rangle, \tag{6.52}$$

then we see that

$$\boxed{U(t, t_0) = e^{-i\Delta t H(N\Delta t)} e^{-i\Delta t H((N-1)\Delta t)} \ldots e^{-i\Delta t H(M\Delta t)}.} \tag{6.53}$$

Since this becomes exact when $\Delta t \to 0$, we conclude that this is *the* general solution of the time-dependent problem. Admittedly, this solution is still quite a mess and part of our effort below will be to clean it up a bit. But if you ever wonder "what is the solution with a time-dependent Hamiltonian?", I recommend you think about (6.53).

It is interesting to note that this operator $U(t, t_0)$ satisfies all properties of its time-independent cousin:

$$U(t_0, t_0) = 1, \tag{6.54}$$

$$U(t, t_1)U(t_1, t_0) = U(t, t_0), \qquad t_0 < t_1 < t, \tag{6.55}$$

$$U(t, t_0)U^\dagger(t, t_0) = 1, \tag{6.56}$$

$$U^\dagger(t, t_0) = U(t_0, t). \tag{6.57}$$

Eq. (6.55) is particularly important, because it shows that even in the time-dependent case the solution can still be broken down in pieces.

The important point that must be remembered concerning Eq. (6.53) is that in general you *cannot* recombine the exponentials since the Hamiltonian at different times may not commute:

$$\text{in general} \quad [H(t), H(t')] \neq 0. \tag{6.58}$$

If this happens to be the case, then the problem is very easy and Eq. (6.53) becomes

$$U(t,t_0) = \exp\Big\{-i\Delta t\sum_{n=M}^{N} H(n\Delta t)\Big\}$$
$$= \exp\Big\{-i\int_{t_0}^{t} H(t')\,\mathrm{d}t'\Big\},$$

where, in the last line, I already took the limit $\Delta t \to 0$ and transformed the sum to an integral.

However, if $H(t)$ does not commute at different times, this solution is incorrect. Instead, we can use a trick to write down the solution in a way that looks formally similar. We define the *time-ordering* operator $\mathcal{T}$ such that, when acting on any set of time-dependent operators, it always puts later times to the left:

$$\mathcal{T}A(t_1)A(t_2) = \begin{cases} A(t_1)A(t_2) & \text{if } t_1 > t_2 \\ A(t_2)A(t_1) & \text{if } t_2 > t_1 \end{cases} \tag{6.59}$$

This time-ordering operator can now be used to combine exponentials. If we recall the Zassenhaus (BCH) formula (**??**):

$$e^{t(A+B)} = e^{tA}e^{tB}e^{-\frac{t^2}{2}[A,B]}e^{\frac{t^3}{3!}(2[B,[A,B]]+[A,[A,B]])}\ldots, \tag{6.60}$$

we see that the combination-recombination of exponentials involves only commutators. Now suppose $t_2 > t_1$. Then

$$\mathcal{T}[A(t_2),B(t_1)] = \mathcal{T}\Big(A(t_2)B(t_1) - B(t_1)A(t_2)\Big) = A(t_2)B(t_1) - A(t_2)B(t_1) = 0.$$

Consequently, if we expand $e^{A(t_2)+B(t_1)}$ and then apply $\mathcal{T}$, the only term that will survive will be $e^{A(t_2)}e^{B(t_1)}$. Hence,

$$e^{A(t_2)}e^{B(t_1)} = \mathcal{T}e^{A(t_2)+B(t_1)}. \tag{6.61}$$

Within the protection of the time-ordering operator, we can freely recombine exponentials.

Using this time-ordering trick we may now recombine all terms in the product (6.53), leading to

$$\boxed{U(t,t_0) = \mathcal{T}\exp\Big\{-i\int_{t_0}^{t} H(t')\,\mathrm{d}t'\Big\},} \tag{6.62}$$

where I already transformed this into an integral. This is the way we usually write the formal solution of a time-dependent problem. The time-ordering operator $\mathcal{T}$ is just a compact way to write down the solution in Eq. (6.53). If you are ever confused about how to operate with it, go back to Eq. (6.53). Finally, let me mention that Eq. (6.62) can also be viewed as the solution of the initial value problem

$$\frac{\mathrm{d}U(t,t_0)}{\mathrm{d}t} = -iH(t)U(t,t_0), \qquad U(t_0,t_0) = 1. \tag{6.63}$$

This may not be so evident from Eq. (6.62), but it is if we substitute Eq. (6.52) into (6.49).

Magnus expansion

We are now in a good point to discuss time-dependent perturbation theory. The scenario is as follows. We start with $H_0 + V$ and move to the interaction picture where the rotating frame Hamiltonian becomes the time-independent operator (6.40). We then try to solve the von Neumann equation for this operator. Or, what is equivalent, we try to find the time-evolution operator $\tilde{U}(t,t_0)$ which, as in (6.63), will be the solution of

$$\frac{\mathrm{d}\tilde{U}(t,t_0)}{\mathrm{d}t} = -i\tilde{H}(t)\tilde{U}(t,t_0), \qquad \tilde{U}(t_0,t_0) = 1. \tag{6.64}$$

There are many ways to do this. Sometimes the perturbation theory is done in terms of states and sometimes it is done in terms of operators (in which case it is called a **Dyson series**).

Here I will try to do it in a slightly different way, using something called a Magnus expansion. Parametrize the time evolution operator as

$$\tilde{U}(t,t_0) = e^{-i\Omega(t,t_0)}, \qquad \Omega(t_0,t_0) = 0, \tag{6.65}$$

where $\Omega(t,t_0)$ is an operator to be determined. To find an equation for it, we first multiply Eq. (6.64) bu $U^\dagger$ on the left, leading to

$$\frac{\mathrm{d}e^{-i\Omega}}{\mathrm{d}t}e^{i\Omega} = -i\tilde{H}(t).$$

Then we use Eq. (6.30) to find

$$\dot{\Omega} - \frac{i}{2}[\Omega,\dot{\Omega}] - \frac{1}{3!}[\Omega,[\Omega,\dot{\Omega}]] + \ldots = \tilde{H}(t), \tag{6.66}$$

which is a really weird equation for $\Omega(t,t_0)$.

We now write this in perturbation-theory-style by assuming that $\tilde{H}(t) \to \epsilon\tilde{H}(t)$ where ϵ is a small parameter. Moreover, we expand Ω as

$$\Omega = \epsilon\Omega_1 + \epsilon^2\Omega_2 + \epsilon^3\Omega_3 + \ldots. \tag{6.67}$$

Substituting in Eq. (6.66) and collecting terms of the same order in ϵ we are then led to a system of equations

$$\dot{\Omega}_1 = \tilde{H}(t), \tag{6.68}$$

$$\dot{\Omega}_2 = \frac{i}{2}[\Omega_1,\dot{\Omega}_1], \tag{6.69}$$

$$\dot{\Omega}_3 = \frac{i}{2}[\Omega_1,\dot{\Omega}_2] + \frac{i}{2}[\Omega_2,\dot{\Omega}_1] + \frac{1}{3!}[\Omega_1,[\Omega_1,\dot{\Omega}_1]]. \tag{6.70}$$

and so on. These can now be solved sequentially, leading to

$$\Omega_1(t) = \int\limits_{t_0}^{t} \mathrm{d}t_1 \; \tilde{H}(t_1), \tag{6.71}$$

$$\Omega_2(t) = -\frac{i}{2}\int\limits_{t_0}^{t} \mathrm{d}t_1 \int\limits_{t_0}^{t_1} \mathrm{d}t_2 \; [\tilde{H}(t_1), \tilde{H}(t_2)], \tag{6.72}$$

$$\Omega_3(t) = -\frac{1}{6}\int\limits_{t_0}^{t} \mathrm{d}t_1 \int\limits_{t_0}^{t_1} \mathrm{d}t_2 \int\limits_{t_0}^{t_2} \mathrm{d}t_3 \; \Big\{[\tilde{H}(t_1), [\tilde{H}(t_2), \tilde{H}(t_3)]] + [\tilde{H}(t_3), [\tilde{H}(t_2), \tilde{H}(t_1)]]\Big\}. \tag{6.73}$$

This is the Magnus expansion. Higher order terms become more and more cumbersome. From this one may obtain the Dyson series expanding Eq. (6.65) in a Taylor series.

It is also important to note that if the Hamiltonian commutes at different times, then the series truncates at the first term. If this were always the case, there would be no need for perturbation theory at all. The need for time-dependent perturbation theory is really a consequence of the non-commutativity of $\tilde{H}$ at different times.

Rotating wave approximation

Consider once again the interaction picture Rabi Hamiltonian (6.41) and let us compute the first order term in the Magnus expansion, Eq. (6.71). We get, assuming $t_0 = 0$,

$$\begin{aligned}\int\limits_{0}^{t} \mathrm{d}t_1 \; \tilde{H}(t_1) = \frac{\lambda}{i(\Omega-\omega_c)}\Big[a\sigma_+(e^{i(\Omega-\omega_c)t} - 1) - a^\dagger\sigma_-(e^{-i(\Omega-\omega_c)t} - 1)\Big] \\ + \frac{\lambda}{i(\Omega+\omega_c)}\Big[a^\dagger\sigma_+(e^{i(\Omega+\omega_c)t} - 1) - a\sigma_-(e^{-i(\Omega+\omega_c)t} - 1)\Big].\end{aligned}$$

The Rotating-wave approximation scenario is now apparent: when we do perturbation theory, the Jaynes-Cummings terms will multiply $\lambda/(\Omega-\omega_c)$ whereas the non-JC terms will contain $\lambda/(\Omega-\omega_c)$. If we are close to resonance ($\Omega \sim \omega_c$) and if λ is small the first term will be very large and the second very small. Consequently, the second term may be neglected.

6.4 Coherent states

Coherent states are a very special set of states which form the basis of continuous variables quantum information. In this section we will discuss some of its basic properties. If you ever need more advanced material, I recommend the paper by K. Cahill and R. Glauber in *Phys. Rev.* **177**, 1857-1881 (1969).

We begin by defining the **displacement operator**

$$\boxed{D(\alpha) = e^{\alpha a^\dagger - \alpha^* a}.} \tag{6.74}$$

where α is an arbitrary complex number and α^* is its complex conjugate. The reason why it is called a "displacement" operator will become clear soon. A coherent state is defined as the action of $D(\alpha)$ into the vacuum state:

$$\boxed{|\alpha\rangle = D(\alpha)|0\rangle.} \tag{6.75}$$

We sometimes say that **"a coherent state is a displaced vacuum"**. This sounds like a typical Star Trek sentence: "Oh no! He displaced the vacuum. Now the entire planet will be annihilated!"

$D(\alpha)$ displaces a and $a^\dagger$

Let us first try to understand why $D(\alpha)$ is called a displacement operator. First, one may verify directly from Eq. (6.74) that

$$D^\dagger(\alpha)D(\alpha) = D(\alpha)D^\dagger(\alpha) = 1 \qquad \text{(it is unitary)}, \tag{6.76}$$

$$D^\dagger(\alpha) = D(-\alpha). \tag{6.77}$$

This means that if you displace by a given α and then displace back by $-\alpha$, you return to where you started. Next I want to compute $D^\dagger(\alpha)aD(\alpha)$. To do that we use the BCH formula (1.70):

$$e^A B e^{-A} = B + [A,B] + \frac{1}{2!}[A,[A,B]] + \frac{1}{3!}[A,[A,[A,B]]] + \dots. \tag{6.78}$$

with $B = a$ and $A = \alpha^* a - \alpha a^\dagger$. Using the commutation relations $[a, a^\dagger] = 1$ we get

$$[\alpha^* a - \alpha a^\dagger, a] = \alpha.$$

But this is a c-number, so that all higher order commutators in the BCH expansion will be zero. We therefore conclude that

$$\boxed{D^\dagger(\alpha)aD(\alpha) = a + \alpha.} \tag{6.79}$$

This is why we call D the displacement operator: it displacements the operator by an amount α. Since $D^\dagger(\alpha) = D(-\alpha)$ it follows that

$$D(\alpha)aD^\dagger(\alpha) = a - \alpha. \tag{6.80}$$

The action on $a^\dagger$ is similar: you just need to take the adjoint: For instance

$$D^\dagger(\alpha)a^\dagger D(\alpha) = a^\dagger + \alpha^*. \tag{6.81}$$

The coherent state is an eigenstate of a

What I want to do now is apply a to the coherent state $|\alpha\rangle$ in Eq. (6.75). Start with Eq. (6.79) and multiply by $D(\alpha)$ on the left. Since D is unitary we get $aD(\alpha) = D(\alpha)(a+\alpha)$. Thus

$$a|\alpha\rangle = aD(\alpha)|0\rangle = D(\alpha)(a+\alpha)|0\rangle = D(\alpha)(\alpha)|0\rangle = \alpha|\alpha\rangle,$$

where I used the fact that $a|0\rangle = 0$. Hence we conclude that **the coherent state is the eigenvector of the annihilation operator**:

$$\boxed{a|\alpha\rangle = \alpha|\alpha\rangle.} \tag{6.82}$$

The annihilation operator is not Hermitian so its eigenvalues do not have to be real. In fact, this equation shows that the eigenvalues of a are all complex numbers.

Alternative way of writing D

It is possible to express D in a different way, which may be more convenient for some computations. Using the Zassenhaus formula (6.60) we see that, if it happens that $[A,B]$ commute with both A and B, then

$$e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}. \tag{6.83}$$

Since $[a,a^\dagger] = 1$, we may write

$$\boxed{D(\alpha) = e^{-|\alpha|^2/2} e^{\alpha a^\dagger} e^{-\alpha^* a} = e^{|\alpha|^2/2} e^{-\alpha^* a} e^{\alpha a^\dagger}.} \tag{6.84}$$

This result is useful because now the exponentials of a and $a^\dagger$ are completely separated.

From this result it follows that

$$\boxed{D(\alpha)D(\beta) = e^{(\beta^*\alpha - \alpha^*\beta)/2} D(\alpha+\beta).} \tag{6.85}$$

This means that if you do two displacements in a sequence, it is almost the same as doing just a single displacement; the only thing you get is a phase factor (the quantity in the exponential is purely imaginary).

Poisson statistics

Let us use Eq. (6.84) to write the coherent state a little differently. Since $a|0\rangle = 0$ it follows that $e^{-\alpha a}|0\rangle = |0\rangle$. Hence we may also write Eq. (6.75) as

$$\boxed{|\alpha\rangle = e^{-|\alpha|^2/2} e^{\alpha a^\dagger}|0\rangle.} \tag{6.86}$$

Now we may expand the exponential and use Eq. (6.14) to write $(a^\dagger)^n|0\rangle$ in terms of the number states. We get

$$|\alpha\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle. \tag{6.87}$$

Thus we find that

$$\langle n|\alpha\rangle = e^{-|\alpha|^2/2}\frac{\alpha^n}{\sqrt{n!}}. \tag{6.88}$$

The probability of finding it in a given state $|n\rangle$, given that it is in a coherent state, is therefore

$$|\langle n|\alpha\rangle|^2 = e^{-|\alpha|^2}\frac{(|\alpha|^2)^n}{n!}. \tag{6.89}$$

This is a Poisson distribution with parameter $\lambda = |\alpha|^2$. The photons in a laser are usually in a coherent state and the Poisson statistics of photon counts can be measured experimentally. If you measure this statistics for thermal light you will find that it is not Poisson (usually it follows a geometric distribution). Hence, Poisson statistics is a signature of coherent states.

Orthogonality

Coherent states are not orthogonal. To figure out the overlap between two coherent states $|\alpha\rangle$ and $|\beta\rangle$ we use Eq. (6.86):

$$\langle\beta|\alpha\rangle = e^{-|\beta|^2/2}e^{-|\alpha|^2/2}\langle 0|e^{\beta^* a}e^{\alpha a^\dagger}|0\rangle.$$

We need to exchange the two operators because we know how a acts on $|0\rangle$ and how $a^\dagger$ acts on $\langle 0|$. To do that we use Eq. (6.83):

$$e^{\beta^* a}e^{\alpha a^\dagger} = e^{\alpha a^\dagger}e^{\beta^* a}e^{\beta^*\alpha}. \tag{6.90}$$

We therefore conclude that

$$\boxed{\langle\beta|\alpha\rangle = \exp\left\{\beta^*\alpha - \frac{|\beta|^2}{2} - \frac{|\alpha|^2}{2}\right\}.} \tag{6.91}$$

The overlap of the two states, squared, can be simplified to read:

$$\boxed{|\langle\beta|\alpha\rangle|^2 = \exp\left\{-|\alpha-\beta|^2\right\}.} \tag{6.92}$$

Hence, the overlap between two coherent states decays *exponentially* with their distance. For large α and β they therefore become approximately orthogonal. Also, as a sanity check, if $\beta = \alpha$ then

$$\langle\alpha|\alpha\rangle = 1, \tag{6.93}$$

which we already knew from Eq. (6.75) and the fact that D is unitary. Coherent states are therefore normalized, but they do *not* form an orthonormal basis. In fact, they form an overcomplete basis in the sense that there are more states than actually needed.

Completeness

Even though the coherent states do not form an orthonormal basis, we can still write down a completeness relation for them. However, it looks a little different:

$$\boxed{\int \frac{\mathrm{d}^2\alpha}{\pi}\, |\alpha\rangle\langle\alpha| = 1.} \tag{6.94}$$

This integral is over the entire complex plane. That is, if $\alpha = x + iy$ then $\mathrm{d}^2\alpha = \mathrm{d}x\,\mathrm{d}y$. This is, therefore, just your old-fashioned integral over two variables. The proof of Eq. (6.94) is a little bit cumbersome. You can find it in Gardiner and Zoller.

Trace of a displacement operator

Due to the orthogonality (6.94), you can also use the coherent state basis to compute traces:

$$\mathrm{tr}(O) = \int \frac{\mathrm{d}^2\alpha}{\pi}\, \langle\alpha|O|\alpha\rangle. \tag{6.95}$$

As an example, let us compute the trace of the displacement operator:

$$\mathrm{tr}\, D(\lambda) = \int \frac{\mathrm{d}^2\alpha}{\pi}\, \langle\alpha|D(\lambda)|\alpha\rangle = \int \frac{\mathrm{d}^2\alpha}{\pi}\, \langle 0|D^\dagger(\alpha)D(\lambda)D(\alpha)|0\rangle.$$

But since $D(\alpha)$ is unitary, it infiltrates everywhere:

$$D^\dagger(\alpha)D(\lambda)D(\alpha) = \exp\Big\{D^\dagger(\alpha)(\lambda a^\dagger - \lambda^* a)D(\alpha)\Big\} = e^{\lambda\alpha^* - \lambda^*\alpha} D(\lambda).$$

Thus we get

$$\mathrm{tr}\, D(\lambda) = \int \frac{\mathrm{d}^2\alpha}{\pi} \;= e^{\lambda\alpha^* - \lambda^*\alpha}\langle 0|D(\lambda)|0\rangle = e^{-|\lambda|^2/2}\int \frac{\mathrm{d}^2\alpha}{\pi}\, e^{\lambda\alpha^* - \lambda^*\alpha} \tag{6.96}$$

where I used the fact that $\langle 0|D(\lambda)|0\rangle = \langle 0|\lambda\rangle = e^{-|\lambda|^2/2}$ [Eq. (6.88)].

The remaining integral is actually an important one. Let us write $\alpha = x + iy$ and $\lambda = u + iv$. Then

$$\lambda\alpha^* - \lambda^*\alpha = 2ixv - 2iuy.$$

Thus

$$\int \frac{\mathrm{d}^2\alpha}{\pi}\, e^{\lambda\alpha^* - \lambda^*\alpha} = \int \mathrm{d}x e^{2ixv} \int \mathrm{d}y e^{-2iuy}$$

But each one is now a Dirac delta

$$\int_{-\infty}^{\infty} \mathrm{d}x e^{ixk} = 2\pi\delta(k).$$

Whence

$$\boxed{\int \frac{\mathrm{d}^2\alpha}{\pi} e^{\lambda\alpha^* - \lambda^*\alpha} = \pi\delta(\lambda).} \tag{6.97}$$

where $\delta(\lambda) = \delta(\mathrm{Re}(\lambda))\delta(\mathrm{Im}(\lambda))$. This integral is therefore nothing but the two-dimensional Fourier transform in terms of the complex variable α.

Substituting this in Eq. (6.96) we finally conclude that

$$\boxed{\mathrm{tr}\, D(\lambda) = \pi\, \delta(\lambda),} \tag{6.98}$$

where I omitted the factor of $e^{-|\lambda|^2/2}$ since the Dirac delta make it irrelevant. Using this and Eq. (6.85) also allows us to write the neat formula

$$\mathrm{tr}\Big[D(\alpha)D^\dagger(\beta)\Big] = \pi\delta(\alpha - \beta). \tag{6.99}$$

This is a sort of orthogonality relation, but between operators.

$D(\alpha)$ as a basis for operators

Due to Eqs. (6.98) and (6.99), it turns out that the displacement operators form a basis for the Hilbert space, in the sense that any operator F may be decomposed as

$$\boxed{F = \int \frac{\mathrm{d}^2\alpha}{\pi} f(\alpha) D^\dagger(\alpha)} \tag{6.100}$$

where

$$f(\alpha) := \mathrm{tr}\Big[F D(\alpha)\Big]. \tag{6.101}$$

This is just like decomposing a state in a basis, but we are actually decomposing an operator.

6.5 The Husimi-Q function

A big part of dealing with continuous variables systems is the idea of *quantum phase space*, similarly to the classical coordinate-momentum phase space in classical mechanics. There are many ways to represent continuous variables in phase space. The three most important are the Husimi-Q function, the Wigner function and the Glauber-Sudarshan P function. Each has its own advantages and disadvantages. Since this chapter is meant to be a first look into this topic, we will focus here on the simplest one of them, the Q function.

The Husimi-Q function is defined as the expectation value of the density matrix in a coherent state

$$\boxed{Q(\alpha^*,\alpha) = \frac{1}{\pi}\langle\alpha|\rho|\alpha\rangle.} \tag{6.102}$$

Here α and α^* are to be interpreted as independent variables. If that confuses you, define $\alpha = x + iy$ and interpret Q as a function of x and y. In fact, following the transformation between $a, a^\dagger$ and the quadrature operators q, p in Eq. (6.3), $x/\sqrt{2}$ represents the position in phase space, whereas $y/\sqrt{2}$ represents the momentum.

Using Eq. (6.95) for the trace in the coherent state basis, we get

$$1 = \mathrm{tr}\,\rho = \int \frac{\mathrm{d}^2\alpha}{\pi}\langle\alpha|\rho|\alpha\rangle.$$

Thus, we conclude that the Husimi Q function is normalized as

$$\int \mathrm{d}^2\alpha\; Q(\alpha^*,\alpha) = 1 \tag{6.103}$$

which resembles the normalization of a probability distribution.

If we know Q we can also use it to compute the expectation value of operators. For instance, since $a|\alpha\rangle = \alpha|\alpha\rangle$ it follows that

$$\langle a\rangle = \mathrm{tr}(\rho a) = \int \frac{\mathrm{d}^2\alpha}{\pi}\,\langle\alpha|\rho a|\alpha\rangle = \int \mathrm{d}^2\alpha\; Q(\alpha,\alpha^*)\alpha,$$

which is intuitive. As another example, recalling that $\langle\alpha|a^\dagger = \langle\alpha|\alpha^*$, we get

$$\langle aa^\dagger\rangle = \mathrm{tr}(a^\dagger\rho a) = \int \frac{\mathrm{d}^2\alpha}{\pi}\,\langle\alpha|a^\dagger\rho a|\alpha\rangle = \int \mathrm{d}^2\alpha\; Q(\alpha,\alpha^*)|\alpha|^2.$$

It is interesting to see here how the ordering of operators play a role. Suppose you want to compute $\langle a^\dagger a\rangle$. Then you should first reorder it as $\langle a^\dagger a\rangle = \langle aa^\dagger\rangle - 1$ and then use the above result for $\langle aa^\dagger\rangle$.

More generally, we may obtain a rule for computing the expectation values of *anti-normally ordered* operators. That is, operators which have all $a^\dagger$s to the right. If this is the case then we can easily write

$$\langle a^k(a^\dagger)^\ell\rangle = \int \mathrm{d}^2\alpha\;\; \alpha^k(\alpha^*)^\ell Q(\alpha^*,\alpha). \tag{6.104}$$

Thus, to compute the expectation value of an arbitrary operator, we should first use the commutation relations to put it in anti-normal order and then use this result.

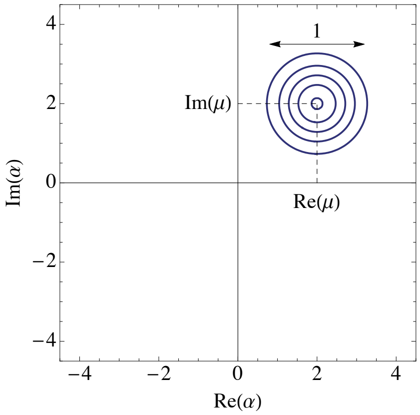

Figure 6.5: Example of the Husimi function (6.106) for $\mu = 2 + 2i$.

The Q function is always non-negative. But not all Q functions correspond to valid states. For instance, $\delta(\alpha)$ is not a valid Husimi function since it would lead to

$$\langle a a^\dagger \rangle = \int \frac{d^2\alpha}{\pi} |\alpha|^2 \delta^2(\alpha) = 0, \tag{6.105}$$

which is impossible since $\langle a a^\dagger\rangle = \langle a^\dagger a\rangle + 1$ and $\langle a^\dagger a\rangle \geq 0$.

Let us now turn to some examples of Q functions.

Example: coherent state

If the state is a coherent state $|\mu\rangle$, then $\rho = |\mu\rangle\langle\mu|$ and we get from (6.92) and (6.102):

$$Q(\alpha^*, \alpha) = \frac{1}{\pi}\langle\alpha|\mu\rangle\langle\mu|\alpha\rangle = \frac{1}{\pi}\exp\Big\{ - |\alpha - \mu|^2\Big\} \tag{6.106}$$

This is a *Gaussian distribution in the complex plane*, centered around μ and with unit variance (see Fig. 6.5). The ground-state of the harmonic oscillator is also a coherent state, but with $\mu = 0$. It will therefore also be a unit-variance Gaussian, but centered at zero. This is why we say *the coherent state is a displaced vacuum*: it has the same distribution, but simply displaced in the complex plane by μ.

Example: Schrödinger cat state

In the context of continuous variables, we sometimes call the superposition

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|\mu\rangle + |-\mu\rangle\Big), \tag{6.107}$$

a Schrödinger cat state. Using Eq. (6.91) we then get

$$Q(\alpha, \alpha^*) = \frac{1}{\pi} e^{-|\alpha-\mu|^2}\Big\{1 + \frac{e^{-2\mu^*\alpha} + e^{-2\mu\alpha^*}}{2}\Big\}. \tag{6.108}$$

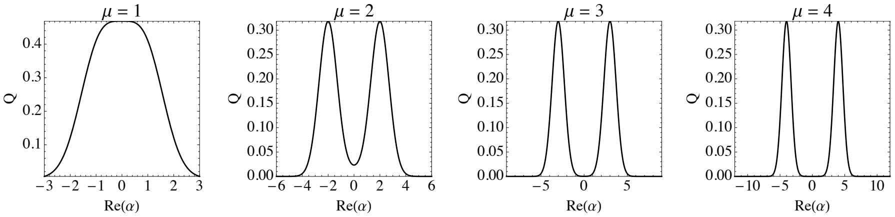

Figure 6.6: Example of the Husimi function (6.108) for a Schrödinger cat state (6.107), assuming μ real. The plots correspond to a cut at $\mathrm{Im}(\alpha) = 0$.

An example of this function is shown in Fig. 6.6. It corresponds to roughly two Gaussians superposed. If μ is small then the two peaks merge into one, but as μ increases they become more distinguishable.

Example: thermal state

Next let us consider a thermal Gibbs state

$$\rho_{\text{th}} = \frac{e^{-\beta\omega a^\dagger a}}{Z}, \tag{6.109}$$

where

$$Z = \operatorname{tr}(e^{-\beta\omega a^\dagger a}) = (1 - e^{-\beta\omega})^{-1}, \tag{6.110}$$

is the partition function. The Husimi function will be

$$Q(\alpha^*, \alpha) = \frac{(1 - e^{-\beta\omega})}{\pi} \sum_{n=0}^{\infty} e^{-\beta\omega n} \langle\alpha|n\rangle\langle n|\alpha\rangle.$$

This is a straightforward and fun calculation, which I will leave for you as an exercise. All you need is the overlap formula (6.88). The result is

$$Q(\alpha^*, \alpha) = \frac{1}{\pi(\bar{n} + 1)} \exp\left\{ -\frac{|\alpha|^2}{\bar{n} + 1} \right\}, \tag{6.111}$$

where

$$\bar{n} = \frac{1}{e^{\beta\omega} - 1}, \tag{6.112}$$

is the Bose-Einstein thermal occupation of the harmonic oscillator. Thus, we see that the thermal state is also a Gaussian distribution, centered at zero but with a variance proportional to $\bar{n} + 1$. At zero temperature we get $\bar{n} = 0$ and we recover the Q function for the vacuum $\rho = |0\rangle\langle 0|$. The width of the Gaussian distribution can be taken as a measure of the fluctuations in the system. At high temperatures $\bar{n}$ becomes large and so does the fluctuations. Thus, in the classical limit we get a big fat Gaussian. But even at $T = 0$ there is still a finite width, which is a consequence of quantum fluctuations.

The two examples above motivate us to consider a **displaced thermal state**. It is defined in terms of the displacement operator (6.74) as

$$\rho = D(\mu)\frac{e^{-\beta\omega a^\dagger a}}{Z}D^\dagger(\mu). \tag{6.113}$$

The corresponding Q function, as you can probably expect, is

$$Q(\alpha^*, \alpha) = \frac{1}{\pi(\bar{n}+1)}\exp\left\{-\frac{|\alpha-\mu|^2}{\bar{n}+1}\right\}, \tag{6.114}$$

which is sort of a mixture of Eqs. (6.106) and (6.111): it represents a thermal Gaussian displaced in the complex plane by an amount μ.

Heterodyne measurements

The Husimi-Q function allows for an interesting interpretation in terms of measurements in the coherent state basis $|\alpha\rangle$, which is called heterodyne measurements. Recall that the basis $|\alpha\rangle$ is not orthonormal and therefore such a measurement is not a projective measurement. Instead, it is a generalized measurement in the same spirit of Sec. 2.7. In particular, please recall Eqs. (2.37)-(2.39). In our case, the set of measurement operators are

$$M_\alpha = \frac{1}{\sqrt{\pi}}|\alpha\rangle\langle\alpha|. \tag{6.115}$$

They are appropriately normalized as

$$\int \mathrm{d}^2\alpha\; M_\alpha^\dagger M_\alpha = \int \frac{\mathrm{d}^2\alpha}{\pi}\;|\alpha\rangle\langle\alpha| = 1,$$

which is nothing but the completeness relation (6.94).

If outcome α is obtained, then the state after the measurement will collapse to $|\alpha\rangle\langle\alpha|$. And the probability of obtaining outcome α is, by Eq. (2.38),

$$\boxed{p_\alpha = \operatorname{tr} M_\alpha \rho M_\alpha^\dagger = \frac{1}{\pi}\langle\alpha|\rho|\alpha\rangle = Q(\alpha, \alpha^*).} \tag{6.116}$$

Thus, we see that the Husimi-Q function is nothing but the probability outcome if we were to perform a heterodyne measurement. This gives a nice interpretation to Q: whenever you see a plot of $Q(\alpha, \alpha^*)$ you can imagine "that is what I would get if I were to measure in the coherent state basis".

6.6 von Neumann's measurement model

In this section I want to use what we learned about continuous variables to discuss a more realistic measurement model. The calculations we are going to do here are a variation of an original proposal given by von Neumann. Suppose we have a system S that has been prepared in some state $|\psi\rangle$ and we wish to measure some observable K in this state. We write the eigenstuff of K as

$$K = \sum_k k|k\rangle\langle k|. \tag{6.117}$$

In order to measure this observable, what we are going to do is couple the system to an ancilla, consisting of a single continuous variable bosonic mode a, according to the interaction Hamiltonian

$$H = igK(a^\dagger - a). \tag{6.118}$$

This Hamiltonian represents a displacement of the bosonic mode which is proportional to the operator K. We could also do the same with $(a + a^\dagger)$ which looks more like a coordinate q. But doing it for $i(a^\dagger - a)$ turns out to be a bit simpler.

We assume the ancila starts in the vacuum so the initial state is

$$|\Phi(0)\rangle_{SA} = |\psi\rangle_S \otimes |0\rangle_A. \tag{6.119}$$

We then compute the time evolution of S+A under the interaction Hamiltonian (6.118). We will not worry here about the free part of the Hamiltonian. Including it would complicate the analysis, but will not lead to any new physics. Our goal then is to compute the state at time t

$$|\Phi(t)\rangle_{SA} = e^{-iHt}|\Phi(0)\rangle_{SA}. \tag{6.120}$$

To evaluate the matrix exponential we expand it in a Taylor series

$$e^{-iHt} = 1 - iHt + \frac{(-i)^2}{2}H^2t^2 + \dots$$

We now note that, using the eigenstuff (6.117), we can write (being a bit sloppy with the $\otimes$):

$$\begin{aligned}
H &= \sum_k |k\rangle\langle k|(igk)(a + a^\dagger), \\
H^2 &= \sum_k |k\rangle\langle k|(igk)^2(a + a^\dagger)^2, \\
&\vdots \\
H^n &= \sum_k |k\rangle\langle k|(igk)^n(a + a^\dagger)^n.
\end{aligned}$$

Thus we may write

$$e^{-iHt} = \sum_k |k\rangle\langle k| e^{gtk(a+a^\dagger)} = \sum_k |k\rangle\langle k| \otimes D(gtk), \tag{6.121}$$

where I introduced here displacement operator $D(\alpha_k) = e^{\alpha_k a^\dagger - \alpha_k^* a}$ [Eq. (6.74)].

It is now easy to apply the evolution operator to the initial state, as in Eq. (6.120). We simply get

$$|\Phi(t)\rangle_{SA} = \sum_k \Big(|k\rangle\langle k| \otimes D(gtk)\Big)\Big(|\psi\rangle_S \otimes |0\rangle_A\Big),$$

or

$$\boxed{|\Phi(t)\rangle_{SA} = \sum_k \big[\langle k|\psi\rangle\big]\, |k\rangle_S \otimes |gtk\rangle_A,} \tag{6.122}$$

where $|gtk\rangle_A = D(gtk)|0\rangle_A$ is the coherent state at position $\alpha = gtk$. This result is quite important. It says that after a time t the combined S+A system will be in an entangled state, corresponding to a superposition of the system being in $|k\rangle$ and the ancilla being in $|gtk\rangle$.

Reduced density matrix of the ancilla

Since the states $|k\rangle$ form an orthonormal basis, the reduced density matrix of the ancilla will be simply

$$\boxed{\rho_A(t) = \text{tr}_S\, |\Phi(t)\rangle\langle\Phi(t)| = \sum_k |\langle k|\psi\rangle|^2 |gtk\rangle\langle gtk|.} \tag{6.123}$$

This is just an incoherent combination of coherent states, with the coherent state $|gtk\rangle$ occurring with probability

$$p_k = |\langle k|\psi\rangle|^2. \tag{6.124}$$

The corresponding Q function will then be simply a sum of terms of the form (6.106):

$$\boxed{Q(\alpha,\alpha^*) = \frac{1}{\pi}\sum_k p_k\, e^{-|\alpha - gtk|^2}.} \tag{6.125}$$

To give an example, suppose our system is a spin 2 particle with dimension $d = 5$ and suppose that the eigenvalues k in Eq. (6.117) are some spin component which can take on the values $k = 2, 1, 0, -1, -2$ [there is nothing special about this example; I'm just trying to give an example that is not based on qubits!]. Suppose also that the state of the system was prepared in

$$|\psi\rangle = \frac{1}{2}\Big\{|2\rangle - |1\rangle - |-1\rangle + |-2\rangle\Big\}, \tag{6.126}$$

where the states here refer to the basis $|k\rangle$ in (6.117). Some examples of the Q function for this state and different values of gt are shown in Fig. 6.7. Remember that the Q function represents a heterodyne detection on the ancilla. These examples show that

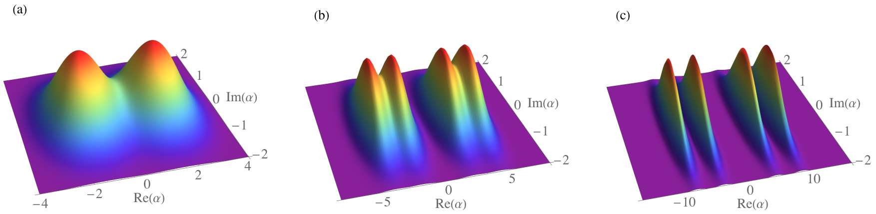

Figure 6.7: Example of the Q function (6.125) computed for the example state (6.126) for different values of gt. Namely (a) 1, (b) 2 and (c) 4.

if gt is small then the different peaks become blurred so such a measurement would not be able to appropriately distinguish between the different peaks. Conversely, as gt gets larger (which means a longer interaction time or a stronger interaction) the peak separation becomes clearer. Thus, the more S and A interact (or, what is equivalent, the more entangled they are) the larger is the amount of information that you can learn about S by performing a heterodyne detection on A.

Reduced density matrix of the system

Next let us compute the reduced density matrix of the system, staring with the composite state (6.122). We get

$$\rho_S(t) = \operatorname{tr}_A |\Phi(t)\rangle\langle\Phi(t)| = \sum_{k,k'} \Big(\langle k|\psi\rangle\langle\psi|k'\rangle\langle gtk|gtk'\rangle\Big)|k\rangle\langle k'|.$$

We can simplify this using the orthogonality relation between coherent states, Eq. (6.91), which gives

$$\langle gtk|gtk'\rangle = \exp\bigg\{-\frac{(gt)^2}{2}(k-k')^2\bigg\}.$$

Thus, the reduced density matrix of S becomes

$$\rho_S(t) = \sum_{k,k'} \rho_{k,k'}(t)|k\rangle\langle k'|, \qquad (6.127)$$

where

$$\boxed{\rho_{k,k'}(t) = \langle k|\psi\rangle\langle\psi|k'\rangle \exp\bigg\{-\frac{(gt)^2}{2}(k-k')^2\bigg\}.} \qquad (6.128)$$

Just as a sanity check, at $t = 0$ we recover the pure state $\rho_S(0) = |\psi\rangle\langle\psi|$.

What is really interesting about Eq. (6.128) is that the diagonal entries of ρ_S in the basis $|k\rangle$ are not effected:

$$\rho_{kk}(t) = \langle k|\psi\rangle\langle\psi|k\rangle = \rho_{k,k}(0). \qquad (6.129)$$

Conversely, the off-diagonal coherences are exponentially damped and if we never turn off the S+A interaction we will eventually end up with

$$\lim_{t\to\infty} \rho_{k,k'}(t) = 0, \qquad k' \neq k. \tag{6.130}$$

Thus, the system initially started in a state $|\psi\rangle$ which was a superposition of the states $|k\rangle$. But, if we allow the system and ancilla to interact for a really long time, the system will end up in a incoherent mixture of states. It is also cool to note how the damping of the coherences is stronger for k and k' which are farther apart.

This analysis shows the emergence of a **preferred basis**. Before we turned on the S+A interaction, the system had no preferred basis. But once that interaction was turned on, the basis of the operator K, which is the operator we chose to couple to the ancila in Eq. (6.118), becomes a preferred basis, in the sense that populations and coherences behave differently in this basis.

Our model also allows us to interpolate between *weak measurements* and *strong measurements*. If gt is small then we perturb the system very little but we also don't learn a lot about it by measuring A. Conversely, if gt is large then we can learn a great deal more, but we also damage the system way more.

Conditional state given measurement outcome

Finally, let us analyze what happens if at time t we perform an actual heterodyne measurement with the operator set M_α in Eq. (6.115). Then if outcome α is obtained, the composite state of S+A will collapse so

$$|\Phi(t)\rangle\langle\Phi(t)| \to \frac{M_\alpha |\Phi(t)\rangle\langle\Phi(t)| M_\alpha^\dagger}{Q(\alpha,\alpha^*)}, \tag{6.131}$$

where I already used Eq. (6.116) to relate the outcome probability p_α with the Husimi function. After the measurement the ancilla will collapse to the coherent state $|\alpha\rangle\langle\alpha|$. Taking the partial trace of Eq. (6.131) over A we then get the reduced density matrix of S, given that the measurement outcome was α. I will leave the details of this calculation to you. The result is

$$\rho_{S|\alpha}(t) = \sum_{k,k'} \rho_{k,k'|\alpha}(t) |k\rangle\langle k'|, \tag{6.132}$$

where

$$\rho_{k,k'|\mu} = \frac{1}{\pi Q(\alpha,\alpha^*)} \langle k|\psi\rangle\langle\psi|k'\rangle\langle\alpha|gtk\rangle\langle gtk'|\alpha\rangle. \tag{6.133}$$

In particular, we can look at the diagonal elements $\rho_{k,k|\alpha}$

$$\rho_{k|\alpha}(t) = \frac{p_k e^{-|\alpha - gtk|^2}}{\sum\limits_{k'} p_{k'} e^{-|\alpha - gtk'|^2}}. \tag{6.134}$$

These quantities represent *the populations in the* $|k\rangle$ *basis, given that the measurement outcome was* α.

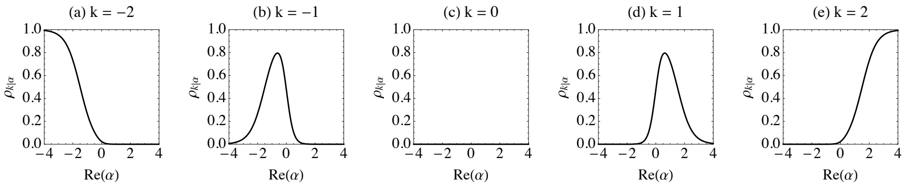

Figure 6.8: The conditional populations in Eq. (6.134) for the example state (6.126) and $gt = 1$.

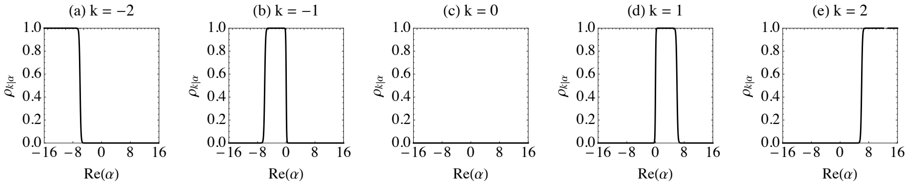

Figure 6.9: Same as Fig. 6.8 but for $gt = 4$.

An example of these conditional populations is shown in Fig. 6.8, which represent $\rho_{k|\alpha}$ for different values of k as a function of Re(α) for the example state (6.126). We can read this as follows. Consider Fig. 6.8(a), which represents $\rho_{-2|\alpha}$. What we see is that if Re(α) $\ll -2$ then it is very likely that the system is found in $k = -2$. Similarly, if Re(α) is around -2, as in Fig. 6.8(b), there is a large probability that the system is found in $k = -1$.

The results in Fig. 6.8 correspond to $gt = 1$ and therefore are not strong measurements. Conversely, in Fig. 6.9) we present the results for $gt = 4$. Now one can see a much shaper distinction of the probabilities. For instance, if Re(α) = 5 then it is almost certain that the system is in $k = 1$, as in Fig. 6.9(d).

6.7 Lindblad dynamics for the quantum harmonic oscillator

We return to the pumped cavity model described in Fig. 6.2. We assume the optical cavity contains only a single mode of radiation a, of frequency ω_c, which is pumped externally by a laser at a frequency ω_p. The Hamiltonian describing this system is given by Eq. (6.22):

$$H = \omega_c a^\dagger a + \epsilon a^\dagger e^{-i\omega_p t} + \epsilon^* a e^{i\omega_p t}. \tag{6.135}$$

In addition to this, we now include also the loss of photons through the semi-transparent mirror. This is modeled by the following master equation

$$\frac{d\rho}{dt} = -i[H,\rho] + D(\rho), \tag{6.136}$$

where $D(\rho)$ is called the **Lindblad dissipator** and is given by

$$D(\rho) = \gamma\Big[a\rho a^\dagger - \frac{1}{2}\{a^\dagger a, \rho\}\Big]. \tag{6.137}$$

Here $\gamma > 0$ is a constant which quantifies the loss rate of the cavity. Recall that the pump term ϵ in Eq. (6.135) was related to the laser power P by $|\epsilon|^2 = \gamma P/\hbar\omega_p$, which therefore depends on γ. This is related to the fact that the mechanism allowing for the photons to get in is the same that allows them to get out, which is the semi-transparent mirror. I should also mention that sometimes Eq. (6.137) is written instead with another constant, $\gamma = 2\kappa$. There is a sort of unspoken rule that if Eq. (6.137) has a 2 in front, the constant should be named κ. If there is no factor of 2, it should be named γ. If you ever want to be mean to a referee, try changing that order.

For qubits the dimension of the Hilbert space is finite so we can describe the master equation by simply solving for the density matrix. Here things are not so easy. Finding a general solution for any density matrix is a more difficult task. Instead, we need to learn alternative ways of dealing with (and understanding) this type of equation.

Before we do anything else, it is important to understand the meaning of the structure of the dissipator, in particular the meaning of a term such as $a\rho a^\dagger$. Suppose at $t = 0$ we prepare the system with certainty in a number state so $\rho(0) = |n\rangle\langle n|$. Then

$$D(|n\rangle\langle n|) = \gamma n\Big\{|n-1\rangle\langle n-1| - |n\rangle\langle n|\Big\}.$$

The first term, which comes from $a\rho a^\dagger$, represents a state with one photon less. This is precisely the idea of a *loss process*. But this process must also preserve probability, which is why we also have another term to compensate. The structure of the dissipator (6.137) represents a very finely tuned equation, where the system looses photons, but does so in such a way that the density matrix remains positive and normalized at all times. We also see from this result that

$$D(|0\rangle\langle 0|) = 0. \tag{6.138}$$

Thus, if you start with zero photons, nothing happens with the dissipator term . We say that the the vacuum is a **fixed point** of the dissipator (it is not necessarily a fixed point of the unitary evolution).

The case of zero pump, $\epsilon = 0$

Let us consider the case $\epsilon = 0$, so that the Hamiltonian (6.135) becomes simply $H = \omega_c a^\dagger a$. This means the photons can never be injected, but only lost. As a consequence, if our intuition is correct, the system should eventually relax to the vacuum. That is, we should expect that

$$\lim_{t\to\infty} \rho(t) = |0\rangle\langle 0|. \tag{6.139}$$

We are going to try to verify this in several ways. The easiest way is to simply verify that if $\rho^* = |0\rangle\langle 0|$ then

$$-i\omega_c[a^\dagger a, \rho^*] + D(\rho^*) = 0,$$

so the vacuum is indeed a steady-state of the equation. If it is unique (it is) and if the system will always converge to it (it will), that is another question.

Next let us look at the populations in the Fock basis

$$p_n = \langle n|\rho|n\rangle. \tag{6.140}$$

They represent the probability of finding the system in the Fock state $|n\rangle$. We can find an equation for $p_n(t)$ by sandwiching Eq. (6.136) in $\langle n| \ldots |n\rangle$. The unitary part turns out to give zero since $|n\rangle$ is an eigenstate of $H = \omega_c a^\dagger a$. As for $\langle n|D(\rho)|n\rangle$, I will leave for you to check that we get

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = \gamma\Big[(n+1)p_{n+1} - np_n\Big]. \tag{6.141}$$

This is called a **Pauli master equation** and is nothing but a **rate equation**, specifying how the population $p_n(t)$ changes with time. Positive terms increase p_n and negative terms decrease it. So the first term in Eq. (6.141) describes the increase in p_n due to populations coming from p_{n+1}. This represents the decays from higher levels. Similarly, the second term in Eq. (6.141) is negative and so describes how p_n decreases due to populations at p_n that are falling down to p_{n-1}.

The steady-state of Eq. (6.141) is obtained by setting $\mathrm{d}p_n/\,\mathrm{d}t = 0$, which gives

$$p_{n+1} = \frac{n}{n+1} p_n, \tag{6.142}$$

In particular, if $n = 0$ we get $p_1 = 0$. Then plugging this in $n = 1$ gives $p_2 = 0$ and so on. Thus, the steady-state correspond to all $p_n = 0$. The only exception is p_0 which, by normalization, must then be $p_0 = 1$.

Evolution of observables

Another useful thing to study is the evolution of observables, such as $\langle a\rangle$, $\langle a^\dagger a\rangle$, etc. Starting from the master equation (6.136), the expectation value of any observables is

$$\frac{\mathrm{d}\langle O\rangle}{\mathrm{d}t} = \mathrm{tr}\left\{O\frac{\mathrm{d}\rho}{\mathrm{d}t}\right\} = -i\,\mathrm{tr}\Big\{O[H,\rho]\Big\} + \mathrm{tr}\Big\{OD(\rho)\Big\}.$$

Rearranging the first term we may write this as

$$\frac{\mathrm{d}O}{\mathrm{d}t} = i\langle[H,O]\rangle + \mathrm{tr}\Big\{OD(\rho)\Big\}. \tag{6.143}$$

The first term is simply Heisenberg's equation (6.48) for the unitary part. What is new is the second term. It is convenient to write this as the trace of ρ times "something", so that we can write this as an expectation value. We can do this using the cyclic property of the trace:

$$\mathrm{tr}\left\{O\Big[a\rho a^\dagger - \frac{1}{2}a^\dagger a\rho - \frac{1}{2}\rho a^\dagger a\Big]\right\} = \langle a^\dagger O a - \frac{1}{2}a^\dagger a O - \frac{1}{2}O a^\dagger a\rangle. \tag{6.144}$$

Using this result for $O = a$ and $O = a^\dagger a$ gives, playing with the algebra a bit,

$$\mathrm{tr}\Big\{a\mathcal{D}(\rho)\Big\} = -\frac{\gamma}{2}\langle a\rangle, \qquad \mathrm{tr}\Big\{a^\dagger a\mathcal{D}(\rho)\Big\} = -\gamma\langle a^\dagger a\rangle. \tag{6.145}$$

Using these results in Eq. (6.143) then gives

$$\frac{\mathrm{d}\langle a\rangle}{\mathrm{d}t} = -(i\omega + \gamma/2)\langle a\rangle, \tag{6.146}$$

$$\frac{\mathrm{d}\langle a^\dagger a\rangle}{\mathrm{d}t} = -\gamma\langle a^\dagger a\rangle. \tag{6.147}$$

Thus, both the first and the second moments will relax exponentially with a rate γ, except that $\langle a \rangle$ will also oscillate:

$$\langle a \rangle_t = e^{-(i\omega + \gamma/2)t} \langle a \rangle_0, \tag{6.148}$$

$$\langle a^\dagger a \rangle_t = e^{-\gamma t} \langle a^\dagger a \rangle_0 \tag{6.149}$$

As $t \to \infty$ the average number of photons $\langle a^\dagger a \rangle$ tends to zero, no matter which state you begin at. Looking at a handful of observables is a powerful way to have an idea about what the density matrix is doing.

Evolution in the presence of a pump

Let us now go back to the full master Eq. (6.136). We can move to the interaction picture exactly as was done in Eq. (6.31), defining

$$\tilde{\rho}_t = S(t) \rho S^\dagger(t), \qquad S(t) = e^{i\omega_p t a^\dagger a}.$$

This transforms the Hamiltonian (6.135) into the detuned time-independent Hamiltonian (6.33):

$$\boxed{\tilde{H} = \Delta a^\dagger a + \epsilon a^\dagger + \epsilon^* a,} \tag{6.150}$$

where $\Delta = \omega_c - \omega_p$ is the detuning. Moreover, I will leave for you as an exercise to check that this does not change in any way the dissipative term. Thus, $\tilde{\rho}$ will evolve according to

$$\frac{\mathrm{d}\tilde{\rho}}{\mathrm{d}t} = -i[\tilde{H}, \tilde{\rho}] + D(\tilde{\rho}). \tag{6.151}$$

To get a feeling of what is going on, let us use Eq. (6.143) to compute the evolution of $\langle a \rangle$. Everything is identical, except for the new pump term that appears. As a result we get

$$\frac{\mathrm{d}\langle a \rangle}{\mathrm{d}t} = -(i\Delta + \gamma/2)\langle a \rangle - i\epsilon. \tag{6.152}$$

As before, $\langle a \rangle$ will evolve as a damped oscillation. However, now it will not tend to zero in the long-time limit, but instead will tend to

$$\langle a \rangle_{\mathrm{ss}} = -\frac{i\epsilon}{i\Delta + \gamma/2}. \tag{6.153}$$

I think this summarizes well the idea of a *pumped cavity*: the steady-state is a competition of how much we pump (unitary term) and how much we drain (the dissipator). Interestingly, the detuning Δ also affects this competition, so for a given ϵ and γ, we get more photons in the cavity if we are at resonance, $\Delta = 0$.

We can also try to ask the more difficult question of what is the density matrix ρ^* in the steady-state. It turns out it is a coherent state set exactly at the value of $\langle a \rangle$:

$$\tilde{\rho}^* = |\alpha\rangle\langle\alpha|, \qquad \alpha = -\frac{i\epsilon}{i\Delta + \gamma/2}. \tag{6.154}$$

One way to check this is to take the coherent state as an ansatz and then try to find what is the value of α which solves Eq. (6.151). The average number of photons will then be

$$\langle a^\dagger a \rangle = |\alpha|^2 = \frac{\epsilon^2}{\Delta^2 + \gamma^2/4}. \tag{6.155}$$

The purpose of this section was to show you a practical use of master equations and open quantum systems. This "cavity loss" dissipator is present in literally every quantum optics setup which involves a cavity. In fact, I know of several papers which sometimes even forget to tell that this dissipator is there, but it always is.

6.8 The spin-boson model

The name spin-boson refers to a single qubit interacting with an infinite number of bosonic modes. Just like with the British royal family, when we talk about this model, there is a good boy and a bad boy. They are:

$$H = \frac{\omega}{2}\sigma_z + \sum_k \Omega_k b_k^\dagger b_k + \sum_k \lambda_k\, \sigma_z(b_k + b_k^\dagger), \qquad \text{(good boy)}, \tag{6.156}$$

$$H = \frac{\omega}{2}\sigma_z + \sum_k \Omega_k b_k^\dagger b_k + \sum_k \lambda_k\, \sigma_x(b_k + b_k^\dagger), \qquad \text{(bad boy)}. \tag{6.157}$$

The fundamental difference between the two models is that in the first the operator appearing in the S-E interaction (σ_z) is the same as the operator in H_S. Consequently, the model (6.156) cannot generate transitions between energy levels (population changes) and, consequently, the most that can happen is environment-induced decoherence. In Eq. (6.157), on the other hand, the operator σ_x is the spin flip and therefore causes population changes. Consequently, it will give rise to an amplitude damping-type of dynamics. In this section we will talk about the good-boy spin-boson model, Eq. (6.156). It has the cool, and quite rare, feature that it can analytically solved.

Exact solution

In this section we find the *exact* solution for $\rho_S(t)$. This is one of the few models for which exact solutions are available, so enjoy it! The starting point is von Neumann's equation (in the Schrödinger picture) for the total density matrix of S+E:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -i[H, \rho], \tag{6.158}$$

where H is the total Hamiltonian (6.156). This is subject to the initial condition

$$\rho(0) = \rho_S(0)\rho_E(0), \qquad \rho_E(0) = \frac{e^{-\beta H_E}}{Z_E}. \tag{6.159}$$

However, now we are interested in exact dynamics so the bath will also evolve in time and the system and bath will become correlated.

The solution of Eq. (6.158) is

$$\rho(t) = e^{-iHt}\rho_S(0)\rho_E(0)e^{iHt}.$$

What we want is the partial trace over the environment

$$\rho_S(t) = \operatorname{tr}_E\Big\{e^{-iHt}\rho_S(0)\rho_E(0)e^{iHt}\Big\}.$$

Let us now divide the total Hamiltonian H as

$$H = H_S + H_0,$$

where $H_S = \frac{\omega}{2}\sigma_z$ and

$$H_0 = \sum_k \Omega_k b_k^\dagger b_k + \sum_k \lambda_k \sigma_z (b_k + b_k^\dagger)$$

The Hamiltonian H_S lives on the qubit space and therefore can be taken out of the partial trace:

$$\rho_S(t) = e^{-iH_S t}\operatorname{tr}_E\Big\{e^{-iH_0 t}\rho_S(0)\rho_E(0)e^{iH_0 t}\Big\}e^{iH_S t}.$$

In this way, we have separated the local unitary dynamics, described by H_S, to the dissipative dynamics described by everything inside the trace. In fact, if you think about it, this whole partial trace is a quantum operation in the spirit of Stinespring's theorem. So for now let us focus on this dissipative part defined by the map

$$\tilde{\rho}_S(t) = \operatorname{tr}_E\Big\{e^{-iH_0 t}\rho_S(0)\rho_E(0)e^{iH_0 t}\Big\}. \tag{6.160}$$

The easiest way to proceed from here is to actually look at the matrix elements of this map in the computational basis. The reason why this is useful is because H_0 is already diagonal in the qubit sector. In fact, we can define

$$H_0|0\rangle = H_0^+|0\rangle, \qquad H_0|1\rangle = H_0^-|1\rangle,$$

where

$$H_0^\pm = \sum_k \Omega_k b_k^\dagger b_k \pm \sum_k \lambda_k (b_k + b_k^\dagger)$$

We then have, for instance

$$\begin{aligned}
\langle 0|\tilde{\rho}_S(t)|0\rangle &= \langle 0|\operatorname{tr}_E\Big\{e^{-iH_0 t}\rho_S(0)\rho_E(0)e^{iH_0 t}\Big\}|0\rangle \\
&= \operatorname{tr}_E\Big\{\langle 0|e^{-iH_0 t}\rho_S(0)\rho_E(0)e^{iH_0 t}|0\rangle\Big\} \\
&= \operatorname{tr}_E\Big\{e^{-iH_0^+ t}\langle 0|\rho_S(0)|0\rangle\rho_E(0)e^{iH_0^+ t}\Big\} \\
&= \langle 0|\rho_S(0)|0\rangle\operatorname{tr}_E\Big\{e^{-iH_0^+ t}\rho_E(0)e^{iH_0^+ t}\Big\}
\end{aligned}$$

This set of steps is important and a bit confusing, so make sure you understand what I am doing. I push the system bra and ket $|0\rangle$ inside the partial trace. But then I know how H_0 acts on it. And after it has acted, H_0^+ will no longer have any components on the qubit space, so we can move $|0\rangle$ through it at will. Finally, when $\langle 0|$ and $|0\rangle$ encounter $\rho_S(0)$, they form a number, which can then be taken outside the partial trace.

But now comes the magic trick: $H_0^\pm$ is an operator that lives only on the environment's Hilbert space. Hence, we are now allowed to use the cyclic property of the trace. This is a useful trick to remember: *if an operator lives on a larger space, cyclic property is forbidden. But if acts only over the space you are tracing, then it becomes allowed again.* And if we do that the two exponentials cancel and we are left with

$$\langle 0|\tilde{\rho}_S(t)|0\rangle = \operatorname{tr}_E\Big\{\langle 0|\rho_S(0)|0\rangle\rho_E(0)\Big\} = \langle 0|\rho_S(0)|0\rangle. \tag{6.161}$$

Thus, as anticipated, we see that the action of the bath does not change the populations (diagonal elements) of ρ_S. A similar argument can of course be used for $\langle 1|\tilde{\rho}_S(t)|1\rangle$ but we don't need to do it because, if $\langle 0|\tilde{\rho}_S(t)|0\rangle$ doesn't change, then $\langle 1|\tilde{\rho}_S(t)|1\rangle$ cannot change also due to normalization.

Next we look at the off-diagonal element

$$\langle 0|\tilde{\rho}_S(t)|1\rangle = \langle 0|\rho_S(0)|1\rangle \operatorname{tr}_E\Big\{e^{-iH_0^+ t}\rho_E(0)e^{iH_0^- t}\Big\}. \tag{6.162}$$

We see now that the exponentials do not cancel, so the result of the trace will not be just $\operatorname{tr}_E \rho_E(0) = 1$. In fact, we can define a general dephasing rate as

$$\boxed{e^{-\Lambda(t)} = \operatorname{tr}_E\Big\{e^{-iH_0^+ t}\rho_E(0)e^{iH_0^- t}\Big\}.} \tag{6.163}$$

Then Eq. (6.162) acquires the more familiar form

$$\langle 0|\tilde{\rho}_S(t)|1\rangle = \langle 0|\rho_S(0)|1\rangle e^{-\Lambda(t)t}. \tag{6.164}$$

Our task has now been reduced to the calculation of the decoherence rate $\Lambda(t)$.

Explicit calculation of the decoherence rate

To compute the trace in Eq. (6.163) we being by noticing that the calculation factors into a product of traces, one for each mode of the environment:

$$\begin{aligned} e^{-\Lambda(t)} &= \prod_k \operatorname{tr}\Big\{e^{-it\left[\Omega_k b_k^\dagger b_k + \lambda_k(b_k + b_k^\dagger)\right]}\rho_k e^{it\left[\Omega_k b_k^\dagger b_k - \lambda_k(b_k + b_k^\dagger)\right]}\Big\} \\ &= \langle e^{it\left[\Omega_k b_k^\dagger b_k - \lambda_k(b_k + b_k^\dagger)\right]} e^{-it\left[\Omega_k b_k^\dagger b_k + \lambda_k(b_k + b_k^\dagger)\right]}\rangle. \end{aligned}$$

where ρ_k is the initial state of mode k of the environment. If we assume the environment is in a thermal state then

$$\rho_k = (1 - e^{-\beta\Omega_k})e^{-\beta\Omega_k b_k^\dagger b_k}.$$

Since the calculations for all modes are equivalent, let us clean up the notation a bit and focus on the quantity

$$B = \langle e^{it\left[\Omega b^\dagger b - \lambda(b+b^\dagger)\right]} e^{-it\left[\Omega b^\dagger b + \lambda(b+b^\dagger)\right]} \rangle \tag{6.165}$$

Computing this is a good exercise on operator algebra.

We will need to recall some definitions of displacement operators $D(\alpha) = e^{\alpha b^\dagger - \alpha^* b}$, discussed in Sec. 6.4. Recall that $D^\dagger(\alpha) b D(\alpha) = b + \alpha$. We can then use this to write

$$\Omega b^\dagger b \pm \lambda(b + b^\dagger) = \Omega\, D^\dagger(\pm\lambda/\Omega)(b^\dagger b) D(\pm\lambda/\Omega) - \frac{\lambda^2}{\Omega}. \tag{6.166}$$

But the displacement operator is unitary, so it can enter or leave exponentials at will. Consequently

$$e^{-it\left[\Omega b^\dagger b + \lambda(b+b^\dagger)\right]} = e^{-it\lambda^2/\Omega} D^\dagger(\lambda/\Omega) e^{-i\Omega t b^\dagger b} D(\lambda/\Omega),$$

with a similar result for the other exponential. Eq. (6.165) then becomes

$$\begin{aligned} B &= \langle D^\dagger(-\lambda/\Omega) e^{i\Omega t b^\dagger b} D(-\lambda/\Omega) D^\dagger(\lambda/\Omega) e^{-i\Omega t b^\dagger b} D(\lambda/\Omega) \rangle \\ &= \langle D(\lambda/\Omega) e^{i\Omega t b^\dagger b} D^\dagger(2\lambda/\Omega) e^{-i\Omega t b^\dagger b} D(\lambda/\Omega) \rangle, \end{aligned}$$

where I used the fact that $D(-\alpha) = D^\dagger(\alpha)$ and that $D(\alpha)D(\alpha) = D(2\alpha)$ (all these properties are described in Sec. 6.4).

In the middle term we infiltrate the exponential inside $D^\dagger(2\lambda/\Omega)$:

$$\begin{aligned} e^{i\Omega t b^\dagger b} D^\dagger(2\lambda/\Omega) e^{-i\Omega t b^\dagger b} &= \exp\left\{ -\frac{2\lambda}{\Omega} e^{i\Omega t b^\dagger b}(b^\dagger - b) e^{-i\Omega t b^\dagger b} \right\} \\ &= \exp\left\{ -\frac{2\lambda}{\Omega}(b^\dagger e^{i\Omega t} - b e^{-i\Omega t}) \right\} \\ &= D^\dagger(2\lambda e^{i\Omega t}/\Omega) \end{aligned}$$

We then arrive at the simpler result:

$$B = \langle D(\lambda/\Omega) D^\dagger(2\lambda e^{i\Omega t}/\Omega) D(\lambda/\Omega) \rangle$$

Finally, we combine the three displacement operators using $D(\alpha)D(\beta) = e^{(\beta^*\alpha - \alpha^*\beta)/2} D(\alpha + \beta)$ [c.f. Eq. (6.85)]. We then finally arrive at

$$\boxed{B = \langle D(\alpha_t) \rangle, \qquad \alpha_t := \frac{2\lambda}{\Omega}(1 - e^{i\Omega t}).} \tag{6.167}$$

This result is somewhat general since it holds for an arbitrary bath initial state.

Next let us specialize it for the case of a thermal state. In this case, it turns out that the trace of a displacement operator in a thermal state is:[3]

$$\boxed{B = \langle D(\alpha_t)\rangle = \exp\bigg\{ - |\alpha_t|^2(\bar{n} + 1/2)\bigg\}, \qquad \text{when} \;\; \rho = (1 - e^{-\beta\Omega})e^{-\beta\Omega b^\dagger b},} \tag{6.168}$$

where $\bar{n} = (e^{\beta\Omega} - 1)^{-1}$.

Decoherence rate for a finite number of modes

Going back now to Eq. (6.163) and reintroducing an index k everywhere, we get

$$e^{-\Lambda(t)} = \prod_k \exp\bigg\{ - |\alpha_{k,t}|^2(\bar{n}(\Omega_k) + 1/2)\bigg\}.$$

Or, taking the log on both sides,

$$\Lambda(t) = \sum_k \frac{4\lambda_k^2}{\Omega_k^2}[1 - \cos(\Omega_k t)] \coth\left(\frac{\Omega_k}{2T}\right), \tag{6.169}$$

where I used the fact that $2\bar{n}(x) + 1 = \coth(x/2T)$ and $|\alpha_t|^2 = \frac{8\lambda^2}{\Omega^2}(1 - \cos\Omega t)$. I think Eq. (6.169) is pretty cool. It is an exact result and therefore holds for an *arbitrary* number of bath modes. This is therefore a good opportunity for us to try to visualize the transition between non-Markovian and Markovian behavior as the number of bath oscillators increases.

For concreteness let us make some assumptions about Ω_k and λ_k. Let us assume that Ω_k varies linearly between 0 and a maximum *cut-off* value Ω_c. That is, if we take a bath of N modes, then we define

$$\Omega_k = \frac{k}{N}\Omega_c, \qquad k = 1, 2, \ldots, N. \tag{6.170}$$

Moreover, let us assume that the coupling constants vary as

$$\lambda_k = \sqrt{\frac{\Omega_k}{N}}. \tag{6.171}$$

---

[3] One way to derive this result is to write it as

$$\langle D(\alpha)\rangle = (1 - e^{-\beta\Omega})\sum_{n=0}^{\infty} e^{-\beta\Omega n}\langle n|D(\alpha)|n\rangle.$$

In K. Cahill and R. Glauber in *Phys. Rev.* **177**, 1857-1881 (1969), they show that $\langle n|D(\alpha)|n\rangle = e^{-|\alpha|^2/2}L_n(|\alpha|^2)$ where $L_n(x)$ are the Laguerre polynomials. The sum in n may then be related to the generating function of Laguerre polynomials:

$$\sum_{n=0}^{\infty} x^n L_n(y) = \frac{1}{1-x}\exp\bigg\{ - \frac{yx}{1-x}\bigg\}.$$

Using this yields, after some simplifications, the result in Eq. (6.168).

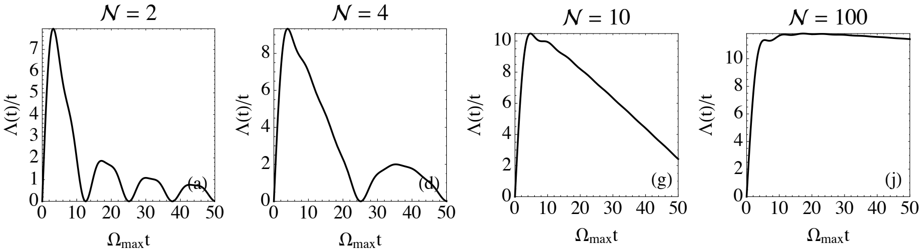

Figure 6.10: The decoherence rate $\Lambda(t)/t$ defined in Eq. (6.169) for different numbers of bath modes N. The parameters Ω_k and λ_k were chosen as in Eqs. (6.170) and (6.171). The temperature was fixed at $T/\Omega_c = 1$.

The logic behind this will be explained below, but essentially it is the condition to obtain what is called an **Ohmic bath**. We also rescale the λ_k with the number of modes since this allows us to compare different values of N.

We present some results for different values of N in Fig. 6.10. As can be seen, if $N = 2$ is small the damping rate is first positive but then goes back all the way to zero at certain points. Having $\Lambda(t) = 0$ means the system didn't dephase at all. This is a signature of non-Markovian behavior. For initial times there is some dephasing. But then information backflows towards the system and it can eventually get back exactly to its initial state when $\Lambda(t) = 0$. As we increase N these backflows start to become more seldom and also occur at larger and larger times. Then, as $N \to \infty$ information never flows back and the dynamics becomes Markovian.

On the other hand, for large N we see that at large times $\Lambda(t)/t$ tends to a constant. This means that the decoherence behaves as $q(t) = q_0 e^{-\Lambda_0 t}$, which is the same result one obtains from the Lindblad equation

$$\frac{d\rho}{dt} = \Lambda\Big[\sigma_z \rho \sigma_z - \rho\Big]. \tag{6.172}$$

But we also see, for instance in the curve with $N = 100$, that for small times there is an adjustment period in which $\Lambda(t)/t$ is not constant. So this means that for very short times there is always some weird stuff going on, even if the bath is infinitely large. The microscopic derivations of master equations don't capture this type of effect because they only take into account a coarse-graining dynamics at large times.

Decoherence rate for an infinite number of modes

The complicated behavior of $\Lambda(t)$ can be clarified if we assume that the number of modes is infinite. In this case we can introduce the spectral density of the bath

$$J(\Omega) = 2\pi \sum_k \lambda_k^2\, \delta(\Omega - \Omega_k),$$

so that Eq. (6.169) becomes

$$\Lambda(t) = \frac{2}{\pi}\int\limits_0^\infty d\Omega\, \frac{J(\Omega)}{\Omega^2}(1-\cos\Omega t)\coth\left(\frac{\Omega}{2T}\right). \tag{6.173}$$

We continue to assume Eqs. (6.170) and (6.171) for Ω_k and λ_k. Since $\lambda_k \sim \sqrt{\Omega_k}$ and $J(\Omega) \sim \lambda_k^2$, we see that these assumptions imply an Ohmic spectral density $J(\Omega) \sim \Omega$. As for the cut-off, we have two choices. One is to assume $J(\Omega) = 0$ when $\Omega > \Omega_c$ (a *hard* cut-off) and the other is to assume that $J(\Omega) \sim e^{-\Omega/\Omega_c}$ (a *soft* cut-off). We shall take the latter. That is, we shall assume that

$$J(\Omega) = A\Omega e^{-\Omega/\Omega_c}, \tag{6.174}$$

where A is some positive dimensionless pre-factor, which we will henceforth set to $A = 1$.

The calculation of the decoherence rate (6.173) now reduces to the following integral:

$$\Lambda(t) = \gamma\int\limits_0^\infty d\Omega\, \frac{e^{-\Omega/\Omega_c}}{\Omega}(1-\cos\Omega t)\coth\left(\frac{\Omega}{2T}\right). \tag{6.175}$$

This integral can actually be played with analytically. You will find this analysis on Sec. 4.2 of Breuer and Petruccione. The result is:

$$\Lambda(t) \simeq \begin{cases} \frac{\Omega_c^2 t^2}{2} & t \ll \Omega_c^{-1} \\ \ln\Omega t & \Omega_c^{-1} \ll t \ll \frac{1}{\pi T} \\ \pi t T & \frac{1}{\pi T} \ll t \end{cases} \tag{6.176}$$

Here Ω_c^{-1} and $1/(\pi T)$ represent characteristic time scales of the problem. The first is a very small time scale (because the cut-off is usually insanely large) and describes the behavior at very short times. Conversely, $1/(\pi T)$ dominates the behavior of the system at the short time scales.

Macroscopic decoherence in multipartite states

We can view the action of the bosons on the spin, Eq. (6.160) as a dephasing channel $\mathcal{E}(\rho_S)$. As we have seen, this channel is such that it does not affect the diagonal elements of ρ_S, only the coherences. We can write the effects of this channel in a convenient way if we move to a spin representation. I will therefore use the mapping

$$|0\rangle \equiv |+1\rangle, \qquad |1\rangle \equiv |-1\rangle.$$

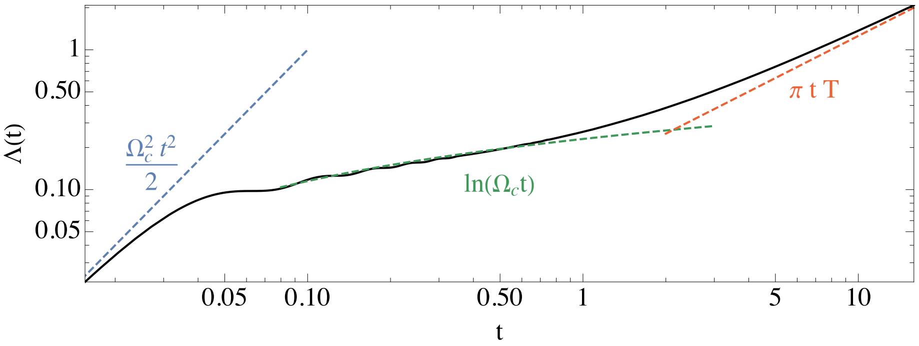

Figure 6.11: The three regimes of the decoherence rate, Eq. (6.176), compared with numerical simulations for $N = 10^4$ bath modes. The other parameters were $\Omega_c = 100$ and $T = 1$.

I will also denote the spin states by a variable $\sigma = \pm 1$. The action of the channel will then be given by

$$\langle +1|\mathcal{E}(\rho_S)|+1\rangle = \langle +1|\rho_S|+1\rangle,$$

$$\langle -1|\mathcal{E}(\rho_S)|-1\rangle = \langle -1|\rho_S|-1\rangle,$$

$$\langle +1|\mathcal{E}(\rho_S)|-1\rangle = \langle +1|\rho_S|-1\rangle e^{-\Lambda(t)},$$

$$\langle -1|\mathcal{E}(\rho_S)|+1\rangle = \langle -1|\rho_S|+1\rangle e^{-\Lambda(t)}.$$

We can write all of this in a single neat formula as

$$\boxed{\langle\sigma|\mathcal{E}(\rho_S)|\sigma'\rangle = \langle\sigma|\rho_S|\sigma'\rangle e^{-\frac{\Lambda(t)}{4}(\sigma-\sigma')^2},} \tag{6.177}$$

where $\sigma, \sigma' = \pm 1$.

Now I want to consider what happens when we have N spins, with each spin coupled to its own individual spin-boson bath. In this case the computational basis will be given by the 2^N kets $|\boldsymbol{\sigma}\rangle = |\sigma_1, \ldots, \sigma_N\rangle$. The logic behind Eq. (6.177) will now continue to apply, but at the level of each individual state. For instance, suppose only spin i was coupled to a spin-boson bath. Then the action of this bath would be

$$\langle\boldsymbol{\sigma}|\mathcal{E}_i(\rho_S)|\boldsymbol{\sigma}'\rangle = \langle\boldsymbol{\sigma}|\rho_S|\boldsymbol{\sigma}'\rangle e^{-\frac{\Lambda(t)}{4}(\sigma_i-\sigma_i')^2}.$$

That is to say, the bath will only cause decoherence in those states of ρ_S which has non-diagonal elements in spin component i. For instance, in the case of two qubits, a state of the form

$$\rho_S = p|+1\rangle\langle+1|+1\otimes|\psi\rangle\langle\psi|_2 + (1-p)|+1\rangle\langle-1|-1\otimes|\psi'\rangle\langle\psi'|_2,$$

would not suffer any decoherence if we applied $\mathcal{E}_1$, but would in general suffer decoherence if we applied $\mathcal{E}_2$, provided $|\psi\rangle$ and $|\psi'\rangle$ have off-diagonal elements.

If we now apply the map to all spins, the effects will simply add up and so we will get

$$\boxed{\langle\boldsymbol{\sigma}|(\mathcal{E}_1\otimes\ldots\otimes\mathcal{E}_N)(\rho_S)|\boldsymbol{\sigma}'\rangle = \langle\boldsymbol{\sigma}|\rho_S|\boldsymbol{\sigma}'\rangle e^{-\frac{\Lambda(t)}{4}\sum\limits_{i=1}^{N}(\sigma_i-\sigma_i')^2}.} \tag{6.178}$$

This equation actually hides a dramatic effect, which explains why quantum effects are so easily washed away in macroscopic systems (like Schrödinger's cat, for instance). To see this, suppose that the system is prepared in a GHZ state,

$$|\psi\rangle = \frac{|0\ldots0\rangle+|1\ldots1\rangle}{\sqrt{2}} := \frac{|\mathbf{0}\rangle+|\mathbf{1}\rangle}{\sqrt{2}}. \tag{6.179}$$

The corresponding density matrix $\rho_S = |\psi\rangle\langle\psi|$ will have four terms. Two of them will not suffer decoherence; namely, $|\mathbf{0}\rangle\langle\mathbf{0}|$ and $|\mathbf{1}\rangle\langle\mathbf{1}|$. But the cross terms $|\mathbf{0}\rangle\langle\mathbf{1}|$ and $|\mathbf{1}\rangle\langle\mathbf{0}|$ will. And according to Eq. (6.178), this decoherence will be proportional to

$$\frac{1}{4}\sum_{i=1}^{N}(\sigma_i-\sigma_i')^2 = N$$

Thus, the evolve state of the system will be

$$\rho_S(t) = \frac{1}{2}\Big\{|\mathbf{0}\rangle\langle\mathbf{0}| + |\mathbf{1}\rangle\langle\mathbf{1}| + |\mathbf{0}\rangle\langle\mathbf{1}|e^{-N\Lambda(t)} + |\mathbf{1}\rangle\langle\mathbf{0}|e^{-N\Lambda(t)}\Big\}. \tag{6.180}$$

Now stop and think about how dramatic this is. Even if Λ is small, we have a factor of N in the exponential. A state such as the GHZ state is very unusual. It is a superposition of all spins up and all spins down. Imagine how weird this would look like in a ferromagnet, for instance. And what Eq. (6.180) is saying is that such a state would decohere exponentially fast with N.