# Chapter 1

# Review of quantum mechanics

Quantum mechanics is all about **states** and **operators**. States represent the instantaneous configuration of your system. You have probably seen them in the form of kets, such as $|\psi\rangle$ and $|i\rangle$, or as wave-functions $\psi(x)$. However, as we will soon learn, the real state in quantum mechanics is specified by an object called a **density matrix**, $\rho$. Density matrices encompass all the information you can have about a physical system, with kets and wave-functions being simply particular cases.

You have also probably seen several examples of operators, such as $H$, $\hat{p}$, $a^\dagger$, $\sigma_z$, etc. *Operators act on states to produce new states.* For instance, the operator $\sigma_x$ flips the 0 and 1's of a qubit, whereas the operator $a^\dagger a$ counts the number of photons in the state. Understanding the action of operators on states 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 Operation** where instead of operators we use *super-operators* (which, you have to admit, sounds cool).

Finally, we have **measurements**. Measurements are also implemented by operators. For instance, that wave-function collapse idea is what we call a *projective measurements* and is implemented by a projection operator. In this course you will also learn about **generalized measurements and POVMs**.

We will actually have a lot to say about measurements in general. Not only are they the least intuitive aspect of quantum mechanics, but they are also the source of all weird effects. If measurements did not exist, quantum mechanics would be quite simple. For many decades, the difficulties concerning the process of measurement were simply swept under the rug. But in the last 4 decades, partially due to experimental advances and fresh new theoretical ideas, this subject has seen a revival of interest. In this course I will adopt the so-called Darwinistic approach: *measurements result from the interaction of a system with its environment*. This interaction is what enables amplification (the process in which quantum signals reach our classical human eyes) and ultimately defines the transition from quantum to classical. But I should probably stop here. I promise we will discuss more about this later.

Before getting our hands dirty, I just want to conclude by saying that, in a simplified view, the above discussion essentially summarizes quantum mechanics, as being formed of three parts: **states, evolutions and measurements**. You have probably seen all three in the old-fashioned way. In this course you will learn about their modern generalizations and how they can be used to construct new technologies. Even though I would love to jump right in, we must start slow. In this chapter I will review most of the linear algebra used in quantum mechanics, together with some results you may have seen before, such as projective measurements and Schrödinger's equation. This will be essential to all that will be discussed in the remaining chapters.

## 1.1 Hilbert spaces and states

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., where the quantity inside the $|\rangle$ is nothing but a *label* to specify which state we are referring to.

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 $\psi_i$ 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.

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

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

Wave-functions are not very useful in this field. 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**. Note that when we refer to a qubit, we don't make any mention to the physical system it represents. In fact, a qubit may represent many physical situations, the two most common being spin 1/2 particles, two-level atoms and the two polarization directions of a photon. A spin 1/2 particle is characterized by spin projections $\uparrow$ and $\downarrow$ in a given direction, so we can label $|0\rangle \equiv |\uparrow\rangle$ and $|1\rangle \equiv |\downarrow\rangle$. Atoms, on the other hand, have very many energy levels. However, sometimes it is reasonable to assume that only the ground state and the first excited state are important, which will be reasonable when the other excited states live too far up the energy ladder. In this case we can make the association $|0\rangle \equiv |g\rangle$, the ground-state, and $|1\rangle \equiv |e\rangle$, the first excited state. Finally, for the polarization of a photon we can call $|0\rangle = |x\rangle$ and $|1\rangle = |y\rangle$, which mean a photon polarized either in the $x$ or $y$ direction. We will play back and forth with these physical representations of a qubit. So let me summarize the main notations:

$$\begin{aligned} |0\rangle &= |\uparrow\rangle = |g\rangle = |x\rangle, \\ |1\rangle &= |\downarrow\rangle = |e\rangle = |y\rangle. \end{aligned} \tag{1.8}$$

An arbitrary state of a qubit may be written as

$$|\psi\rangle = a|0\rangle + b|1\rangle = \begin{pmatrix} a \\ b \end{pmatrix}, \tag{1.9}$$

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

$$|a|^2 + |b|^2 = 1 \tag{1.10}$$

A convenient way to parametrize $a$ and $b$ is as

$$a = \cos(\theta/2), \qquad b = e^{i\phi}\sin(\theta/2), \tag{1.11}$$

where $\theta$ and $\phi$ 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 (1.10), 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

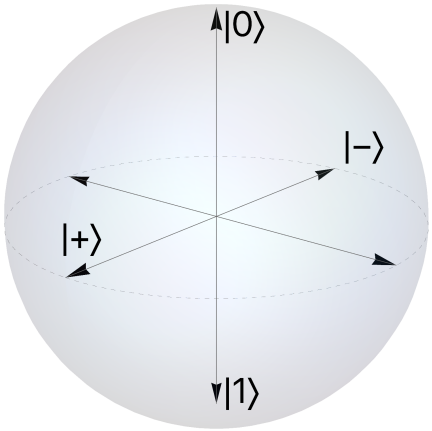


**Figure 1.1:** Example of Bloch's sphere which maps the general state of a qubit into a sphere of unit radius.

representation, the state $|0\rangle$ is the north pole, whereas $|1\rangle$ is the south pole. In this figure I also highlight two other states which appear often, called $|\pm\rangle$. They are defined as

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}. \tag{1.12}$$

In terms of the angles $\theta$ and $\phi$ in Eq. (1.11), this corresponds to $\theta = \pi/2$ and $\phi = 0, \pi$. Thus, these states lie in the *equator*, as show in Fig. 1.1. Also, I should mention a possible source of confusion between the states $|\pm\rangle$ and the up-down states of spin 1/2 particles. I will discuss a way to lift this confusion below, when we talk about Pauli matrices.

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

$$|\psi\rangle\langle\phi| = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c^* & d^* \end{pmatrix} = \begin{pmatrix} ac^* & ad^* \\ bc^* & bd^* \end{pmatrix}. \tag{1.13}$$

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

$$\psi_i = \langle i|\psi\rangle. \tag{1.14}$$

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

$$\boxed{\sum_i |i\rangle\langle i| = 1 = \mathbb{I}} \tag{1.15}$$

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. (1.15) I wrote both the symbol $\mathbb{I}$, which stands for the identity matrix, and the number 1. Using the same symbol for a matrix and a number can feel strange sometimes. The point is that the identity matrix and the number 1 satisfy exactly the same properties and therefore it is not necessary to distinguish between the two.

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 (1.12) 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 (1.15) 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. (1.15) 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

$$P_i = |i\rangle\langle i|. \tag{1.16}$$

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

$$P_i^2 = P_i, \qquad P_iP_j = 0 \quad \text{if } i \neq j, \tag{1.17}$$

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 (1.15) 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:

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

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:

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

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. (1.18) with $\langle j|$, which gives

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

However, I should mention that using the inner product is not strictly necessary to determine matrix elements. It is also possible to define matrix elements for operators acting on vector spaces that are not equipped with inner products. After all, we only need the list of results in Eq. (1.18).

We can also use the completeness (1.15) twice to write

$$A = 1A1 = \sum_{i,j} |i\rangle\langle i|A|j\rangle\langle j| = \sum_{i,j} A_{i,j}|i\rangle\langle j|. \tag{1.20}$$

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,

$$A = \begin{pmatrix} A_{0,0} & A_{0,1} \\ A_{1,0} & A_{1,1} \end{pmatrix} \tag{1.21}$$

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**

$$[A, B] = AB - BA. \tag{1.22}$$

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:

$$[AB, C] = A[B, C] + [A, C]B \tag{1.23}$$

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 every once in a while, 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,

$$A|\lambda = \lambda|\lambda\rangle. \tag{1.24}$$

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

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

$$(A^\dagger)_{i,j} = A^*_{j,i} \tag{1.25}$$

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) eigenvalues.

2. The eigenvalues are always real.

3. The 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 these 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. (1.24) by $\langle\lambda|$, which gives

$$\langle\lambda|A|\lambda\rangle = \lambda. \tag{1.26}$$

Because of the relation (1.25), it now follows for any state that,

$$\langle\psi|A|\phi\rangle = \langle\phi|A^\dagger|\psi\rangle^*. \tag{1.27}$$

Taking the complex conjugate of Eq. (1.26) 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. This result also shows that when $A$ is not Hermitian, if $\lambda$ happens to be an eigenvalue, then $\lambda^*$ will also be an eigenvalue.

Since the eigenvectors $|\lambda\rangle$ form a basis, we can decompose an operator $A$ as in (1.20), but using the basis $\lambda$. We then get

$$A = \sum_\lambda \lambda|\lambda\rangle\langle\lambda|. \tag{1.28}$$

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 unitary matrix $U$ is one that satisfies:

$$UU^\dagger = U^\dagger U = 1, \tag{1.29}$$

where, as above, here 1 means the identity matrix. 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. (1.24)]. Given the set of $d$ eigenvectors $|\lambda_i\rangle$, construct a matrix where each column is an eigenvector:

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

Then

$$U^\dagger = \begin{pmatrix} \dots & \dots & \langle\lambda_0| & \dots & \dots \\ \dots & \dots & \langle\lambda_1| & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \langle\lambda_{d-1}| & \dots & \dots \end{pmatrix} \tag{1.31}$$

But since $A$ is Hermitian, the eigenvectors form an orthonormal basis, $\langle\lambda_i|\lambda_j\rangle = \delta_{i,j}$. You may then verify that this $U$ satisfies (1.29). That is, it is unitary.

To finish the diagonalization procedure, let us also define a diagonal matrix containing the eigenvalues:

$$\Lambda = \mathrm{diag}(\lambda_0, \lambda_1, \dots, \lambda_{d-1}) \tag{1.32}$$

Then, I will leave for you to check that the matrix $A$ in (1.28) may be written as

$$A = U\Lambda U^\dagger \tag{1.33}$$

Thus, we see that any Hermitian matrix may be diagonalized by a Unitary transformation. That is to say, there is always a "rotation" that makes $A$ diagonal. The eigenvector basis $|\lambda_i\rangle$ is the "rotated" basis, where $A$ is diagonal.

## 1.7 Projective measurements and expectation values

As you know, in quantum mechanics measuring a system causes the wavefunction to collapse. The basic measurement describing this (and which we will later generalize) is called a projective measurement. It can be postulated in two ways, either as *measuring in a basis* or *measuring an observable*. Both are

actually equivalent. 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

$$p_i = |\langle i|\psi\rangle|^2 \tag{1.34}$$

Moreover, if the system was found in state $|i\rangle$, then due to the action of the measurement its state has collapsed to the state $|i\rangle$. That is, the measurement transforms the state as $|\psi\rangle \to |i\rangle$. 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 (1.34) 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 $\lambda_i$. 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 $\lambda_i$. 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 $\lambda_i$, 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

$$\langle A\rangle := \sum_i \lambda_i p_i \tag{1.35}$$

I will leave for you to show that using Eq. (1.34) we may also write this as

$$\langle A\rangle := \langle\psi|A|\psi\rangle \tag{1.36}$$

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 (1.34) become

$$p_i = \langle\psi|P_i|\psi\rangle. \tag{1.37}$$

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

$$\boxed{\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \tag{1.38}$$

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

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

The operators $\sigma_x$ and $\sigma_y$, on the other hand, flip the qubit. For instance,

$$\sigma_x|0\rangle = |1\rangle, \qquad \sigma_x|1\rangle = |0\rangle. \tag{1.40}$$

The action of $\sigma_y$ 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**:

$$\boxed{\sigma_+ = |0\rangle\langle 1| = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \sigma_- = |1\rangle\langle 0| = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}} \tag{1.41}$$

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

$$\sigma_x = \sigma_+ + \sigma_- \qquad \text{and} \qquad \sigma_y = -i(\sigma_+ - \sigma_-) \tag{1.42}$$

or

$$\sigma_\pm = \frac{\sigma_x \pm i\sigma_y}{2} \tag{1.43}$$

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

$$\sigma_+|1\rangle = |0\rangle, \qquad \text{and} \qquad \sigma_-|0\rangle = |1\rangle \tag{1.44}$$

This confusion is partially my fault since I defined $|0\rangle = |\uparrow\rangle$ and $|1\rangle = |\downarrow\rangle$. In terms of $\uparrow$ and $\downarrow$ they make sense: the operator $\sigma_-$ lowers the spin value whereas $\sigma_+$ raises it.

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 the theory of angular momentum, where the Pauli matrices appear as the spin operators for spin 1/2 particles. As we will see, this will allow us to make nice a connection with Bloch's sphere. The commutation relations between the Pauli matrices are

$$[\sigma_i, \sigma_j] = 2i\epsilon_{i,j,k}\sigma_k, \tag{1.45}$$

which is the angular momentum algebra, except for the factor of 2. Based on our little table (1.8), we then see that $|0\rangle = |\uparrow\rangle$ and $|1\rangle = |\downarrow\rangle$ are the eigenvectors of $\sigma_z$, with eigenvalues $+1$ and $-1$ respectively. The states $|\pm\rangle$ in Eq. (1.12) are then the eigenstates of $\sigma_x$, also with eigenvalues $\pm 1$. To avoid the confusion a good notation is to call the eigenstates of $\sigma_z$ as $|z_\pm\rangle$ and those of $\sigma_x$ as $|x_\pm\rangle$. That is, $|0\rangle = |z_+\rangle$, $|1\rangle = |z_-\rangle$ and $|\pm\rangle = |x_\pm\rangle$.

As mentioned, the operator $\sigma_i$ is the spin operator at direction $i$. Of course, the orientation of $\mathbb{R}^3$ is a matter of choice, but once we choose a coordinate system, we can then define 3 independent spin operators, one for each of the orthogonal directions. We can also define spin operators in an arbitrary orientation in space. Such an orientation can be defined by a unit vector in spherical coordinates

$$\boldsymbol{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{1.46}$$

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

$$\sigma_{\boldsymbol{n}} = \boldsymbol{\sigma}\cdot\boldsymbol{n} = \sigma_x n_x + \sigma_y n_y + \sigma_z n_z \tag{1.47}$$

Please take a second to check that we can recover $\sigma_{x,y,z}$ just by taking appropriate choices of $\theta$ and $\phi$. In terms of the parametrization (1.46) this spin operator becomes

$$\sigma_{\boldsymbol{n}} = \begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix} = \begin{pmatrix} \cos\theta & e^{-i\phi}\sin\theta \\ e^{i\phi}\sin\theta & -\cos\theta \end{pmatrix} \tag{1.48}$$

I will leave for you to compute the eigenvalues and eigenvectors of this operator. The eigenvalues are $\pm 1$, 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 $\pm 1$. As for the eigenvectors, they are

$$|\boldsymbol{n}_+\rangle = \begin{pmatrix} e^{-i\phi/2}\cos\frac{\theta}{2} \\ e^{i\phi/2}\sin\frac{\theta}{2} \end{pmatrix}, \qquad |\boldsymbol{n}_-\rangle = \begin{pmatrix} -e^{-i\phi/2}\sin\frac{\theta}{2} \\ e^{i\phi/2}\cos\frac{\theta}{2} \end{pmatrix} \tag{1.49}$$

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 (1.11), 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

$$\langle\sigma_x\rangle = \sin\theta\cos\phi, \qquad \langle\sigma_y\rangle = \sin\theta\sin\phi, \qquad \langle\sigma_z\rangle = \cos\theta \tag{1.50}$$

Thus, the average of $\sigma_i$ 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. (1.46), and showed how to write the corresponding state in $\mathbb{C}^2$, Eq. (1.49).

To finish, let us also write the diagonalization of $\sigma_{\boldsymbol{n}}$ in the form of Eq. (1.33). To do that, we construct a matrix whose columns are the eigenvectors $|\boldsymbol{n}_+\rangle$ and $|\boldsymbol{n}_-\rangle$. This matrix is then

$$G = \begin{pmatrix} e^{-i\phi/2}\cos\frac{\theta}{2} & -e^{-i\phi/2}\sin\frac{\theta}{2} \\ e^{i\phi/2}\sin\frac{\theta}{2} & e^{i\phi/2}\cos\frac{\theta}{2} \end{pmatrix} \tag{1.51}$$

The diagonal matrix $\Lambda$ in Eq. (1.33) is the matrix containing the eigenvalues $\pm 1$. Hence it is precisely $\sigma_z$. Thus, we conclude that

$$\sigma_{\boldsymbol{n}} = G\sigma_z G^\dagger \tag{1.52}$$

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

## 1.9 General two-level systems

As we mentioned above, two-state systems appear all the time. And when writing operators for these systems, it is always convenient to express them in terms of Pauli matrices $\sigma_x$, $\sigma_y$, $\sigma_z$ and $\sigma_0 = 1$ (the identity matrix), which can be done for any $2 \times 2$ matrix. We can write this in an organized way as

$$A = a_0 + \boldsymbol{a}\cdot\boldsymbol{\sigma}, \tag{1.53}$$

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

$$A = a_0 + a(\boldsymbol{n}\cdot\boldsymbol{\sigma}) \tag{1.54}$$

Now suppose we wish to find the eigenvalues and eigenvectors of $A$. The eigenvalues are always easy, but the eigenvectors can become somewhat ugly, even in this $2 \times 2$ case. Writing in terms of Pauli matrices makes this more organized. For the eigenvalues, the following silly properties are 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. (1.54), we then see that

$$\text{eigs}(A) = a_0 \pm a \tag{1.55}$$

As for the eigenvectors, they will be given precisely by Eq. (1.49), where the angles $\theta$ and $\phi$ are defined in terms of the unit vector $\boldsymbol{n} = \boldsymbol{a}/a$. Thus, we finally conclude that any $2 \times 2$ matrix may be diagonalized as

$$A = G(a_0 + a\sigma_z)G^\dagger \tag{1.56}$$

This gives an elegant way of writing the eigenvectors of $2 \times 2$ matrices.

## 1.10 Functions of operators

Let $A$ be some Hermitian operator, decomposed as in Eq. (1.28):

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

Now let us compute $A^2$. 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 $\lambda_i^2$, whereas the eigenvectors are the same as those of $A$. Of course, this is also true for $A^3$ or any other power. Now let $f(x)$ be an arbitrary function which can be expanded in a Taylor series, $f(x) = \sum_n c_n x^n$. We can always define the action of this function on operators, instead of numbers, by assuming that the same Taylor series holds for the operators. That is, we define

$$f(A) := \sum_n c_n A^n \tag{1.57}$$

If we now write $A$ in diagonal form, we then see that

$$f(A) = \sum_i f(\lambda_i) |\lambda_i\rangle\langle\lambda_i| \tag{1.58}$$

This is a very useful formula for computing functions of operators.

We can also derive this formula for the case when $A$ is diagonalized as in Eq. (1.33): $A = U\Lambda U^\dagger$. Then, since $UU^\dagger = U^\dagger U = 1$, it follows that $A^2 = U\Lambda^2 U^\dagger$ and so on. By writing $A$ like this, we can now apply any function we want, by simply applying the function to the corresponding eigenvalues:

$$f(A) = U f(\Lambda) U^\dagger \tag{1.59}$$

Since $\Lambda$ is diagonal, the action of $f$ on $\Lambda$ is equivalent to applying $f$ to each diagonal entry.

The most important example is by far the exponential of an operator, defined as

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots, \tag{1.60}$$

Using our two basic formulas (1.58) and (1.59) we then get

$$e^A = \sum_i e^{\lambda_i} |\lambda_i\rangle\langle\lambda_i| = U e^\Lambda U^\dagger \tag{1.61}$$

Another useful example is the inverse:

$$A^{-1} = U\Lambda^{-1} U^\dagger \tag{1.62}$$

To practice, let us compute the exponential of some Pauli operators. We start with $\sigma_z$. Since it is diagonal, we simply exponentiate the entries:

$$e^{i\alpha\sigma_z} = \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix}$$

Next we do the same for $\sigma_x$. The eigenvectors of $\sigma_x$ are the $|\pm\rangle$ states in Eq. (1.12). Thus

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

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. (1.60). Collecting the terms proportional to $\sigma_{\boldsymbol{n}}$ and $\sigma_{\boldsymbol{n}}^2 = 1$ we get:

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

Thus, we readily see that

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

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 $\sigma_x$, we recover Eq. (1.63). It is interesting to point out that nowhere did we use the fact that the matrix was $2 \times 2$. 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. (1.49). 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 $\theta$ and then around the $z$ axis by an angle $\phi$ (take a second to imagine how this works in your head). Thus, one would expect that

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

I will leave for you to check that this is indeed Eq. (1.49). 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. (1.51) can also be shown to be

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

The exponential of an operator is defined by means of the Taylor series (1.60). 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:

$$e^{A+B} \neq e^A e^B. \quad (1.67)$$

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 $e^{i\alpha\sigma_{\boldsymbol{n}}}$ 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 the result of Dr. 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

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

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.

Finally, I wanna mention a trick that is very useful when dealing with general functions of operators. Let $A$ be some operator and define $B = UAU^\dagger$, where $U$ is unitary. Then $B^2 = UA^2U^\dagger$ and etc. Consequently, when we apply a unitary sandwich to any function $f(A)$, we can infiltrate the unitary inside the function:

$$Uf(A)U^\dagger = f(UAU^\dagger). \tag{1.71}$$

This is a little bit more general than (1.59), in which $\Lambda$ was diagonal. But the idea is exactly the same. For instance, with Eq. (1.52) in mind, we can write

$$e^{i\alpha\sigma_{\boldsymbol{n}}} = Ge^{i\alpha\sigma_z}G^\dagger$$

## 1.11 The Trace

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

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

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

$$\mathrm{tr}(A) = \sum_i \langle i|A|i\rangle = \sum_a \langle a|A|a\rangle. \tag{1.73}$$

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

$$\boxed{\mathrm{tr}(A) = \sum_i \lambda_i = \text{sum of all eigenvalues of } A}. \tag{1.74}$$

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

$$\mathrm{tr}(AB) = \mathrm{tr}(BA). \tag{1.75}$$

I will leave it for you to demonstrate this. You can do it, as with all demonstrations in quantum mechanics, by inserting a convenient completeness relation in the middle of $AB$. Using the cyclic property (1.75) you can also move around an arbitrary number of operators, but only in cyclic permutations. For instance:

$$\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA). \tag{1.76}$$

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

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

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 $|\psi\rangle\langle\phi|$:

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

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

$$\boxed{\text{tr}(|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle.} \tag{1.77}$$

Notice how this follows the same logic as Eq. (1.75), so you can pretend you just used the cyclic property. This formula turns out to be extremely useful, so it is definitely worth remembering.

## 1.12 Schrödinger's equation

So far nothing has been said about how states evolve in time. The equation governing the time evolution is called Schödinger's equation. This equation cannot be derived from first principles. It is a postulate of quantum mechanics. Interestingly, however, we don't need to postulate the equation itself. Instead, all we need to postulate is that the transformation caused by the time evolution is 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

$$|\psi(t)\rangle = U(t, t_0)|\psi(t_0)\rangle, \tag{1.78}$$

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 the assumption of linearity, we also have that states must remain normalized. That is, they must always satisfy $\langle\psi|\psi\rangle = 1$ at all times. Looking at Eq. (1.78), we see that this will only be true when the matrix $U(t, t_0)$ is unitary. Hence, we conclude that *time evolution must be described by a unitary matrix.*

Eq. (1.78) 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 $\Delta t$, which to first order can be written as

$$U(t + \Delta t, t) \simeq 1 - i\Delta t H(t) \tag{1.79}$$

where $H(t)$ is some operator which, as you of course know, is called the **Hamiltonian** of your system. The reason why I put the $i$ in front is because then $H$ is 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. (1.79) in Eq. (1.78), dividing by $\Delta t$ and then taking the limit $\Delta t \to 0$ we get

$$\partial_t|\psi(t)\rangle = -iH(t)|\psi(t)\rangle \tag{1.80}$$

which is Schrödinger's equation.

What we have therefore learned is that, once we postulate normalization and linearity, the evolution of a physical system *must* be given by an equation of the form (1.80), 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. We will explore several Hamiltonians along the way.

If the Hamiltonian is time-independent, then the solution of Eq. (1.80) is given by the time-evolution operator

$$U(t, t_0) = e^{-iH(t-t_0)}. \tag{1.81}$$

Even when the Hamiltonian is time-dependent, it is also possible to write a solution that looks like this, but we need to introduce something called the time-ordering operator. We will discuss this later. Eq. (1.81) also has an interesting interpretation concerning the quantization of classical mechanics. When a unitary is written like the exponential of something, we say the quantity in the exponent is the *generator* of that transformation. Thus, **the Hamiltonian is the generator of time-translations**. According to Nöether's theorem in classical mechanics, to every symmetry there is a corresponding conserved quantity. Thus, for instance, when a system is invariant under time translations (i.e. has a time-independent Hamiltonian) then energy is a conserved quantity. In quantum mechanics, the conserved quantity is promoted to an operator and becomes the generator of the symmetry.

To take another example, we know that if a classical system is invariant under rotations, the angular momentum is conserved. Consequently, in the quantum theory, angular momentum will be promoted to an operator and will become the generator of translations. Indeed, as we have already seen, $e^{-i\phi\sigma_z/2}$ is the operator that rotates a ket around the $z$ axis by an angle $\phi$.

Next let us define the eigenstuff of the Hamiltonian as

$$H|n\rangle = E_n|n\rangle. \tag{1.82}$$

Then, using the tricks of Sec. 1.10, we may write the time-evolution operator in Eq. (1.81) as

$$U(t, t_0) = \sum_n e^{-iE_n(t-t')}|n\rangle\langle n|. \tag{1.83}$$

An arbitrary initial state $|\psi_0\rangle$ may always be decomposed in the eigenbasis $|n\rangle$ as $|\psi_0\rangle = \sum_n \psi_n|n\rangle$. Then, the time-evolved state will be

$$|\psi_t\rangle = \sum_n e^{-iE_n(t-t_0)}\psi_n|n\rangle \tag{1.84}$$

Each component in the eigenbasis of the Hamiltonian simply 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 oscillate back and forth forever.

## 1.13 The Schrödinger Lagrangian

It is possible to cast Schrödinger's equation as a consequence of the **principle of least action**, similar to what we do in classical mechanics. This is fun because it formulates quantum mechanics as a classical theory, as weird as that may sound. There is no particular reason why I will introduce this idea here. I just think it is beautiful and I wanted to share it with you.

Let us start with a brief review of classical mechanics. Consider a system described by a set of generalized coordinates $q_i$ and characterized by a Lagrangian $L(q_i, \partial_t q_i)$. The action is defined as

$$S = \int_{t_1}^{t_2} L(q_i, \partial_t q_i)\, \mathrm{d}t. \tag{1.85}$$

The motion of the system is then generated by the principle of least action; ie, by requiring that the actual path should be an extremum of $S$. We can find the equations of motion (the Euler-Lagrange equations) by performing a tiny variation in $S$ and requiring that $\delta S = 0$ (which is the condition on any extremum point; maximum or minimum). To do that we write $q_i \to q_i + \eta_i$, where $\eta_i(t)$ is supposed to be an infinitesimal distortion of the original trajectory. We then compute

$$\begin{aligned}
\delta S &= S[q_i(t) + \eta_i(t)] - S[q_i(t)] \\
&= \int_{t_1}^{t_2} \mathrm{d}t \sum_i \left\{ \frac{\partial L}{\partial q_i}\eta_i + \frac{\partial L}{\partial(\partial_t q_i)}\partial_t \eta_i \right\} \\
&= \int_{t_1}^{t_2} \mathrm{d}t \sum_i \left\{ \frac{\partial L}{\partial q_i} - \partial_t\left( \frac{\partial L}{\partial(\partial_t q_i)} \right) \right\} \eta_i.
\end{aligned}$$

where, in the last line, I integrated by parts the second term. Setting each term proportional to $\eta_i$ to zero then gives us the **Euler-Lagrange equations**

$$\boxed{\frac{\partial L}{\partial q_i} - \partial_t\left( \frac{\partial L}{\partial(\partial_t q_i)} \right) = 0.} \tag{1.86}$$

The example you are probably mostly familiar with is the case when

$$L = \frac{1}{2} m(\partial_t q)^2 - V(q), \tag{1.87}$$

with $V(q)$ being some potential. In this case Eq. (1.86) gives Newton's law

$$m\partial_t^2 q = -\frac{\partial V}{\partial q}. \tag{1.88}$$

Another example, which you may not have seen before, but which will be interesting for us, is the case when we write $L$ with both the position $q$ and the momenta $p$ as generalized coordinates; , ie $L(q, \partial_t q, p, \partial_t p)$. For instance,

$$L = p\partial_t q - H(q, p), \tag{1.89}$$

where $H$ is the Hamiltonian function. In this case there will be two Euler-Lagrange equations for the coordinates $q$ and $p$:

$$\frac{\partial L}{\partial q} - \partial_t\left(\frac{\partial L}{\partial(\partial_t q)}\right) = -\frac{\partial H}{\partial q} - \partial_t p = 0$$
$$\frac{\partial L}{\partial p} - \partial_t\left(\frac{\partial L}{\partial(\partial_t p)}\right) = \partial_t q - \frac{\partial H}{\partial p} = 0.$$

Rearranging, this gives us Hamilton's equations

$$\partial_t p = -\frac{\partial H}{\partial q}, \qquad \partial_t q = \frac{\partial H}{\partial p}. \tag{1.90}$$

Another thing we will need is the **conjugated momentum** $\pi_i$ associated to a generalized coordinate $q_i$. It is always defined as

$$\pi_i = \frac{\partial L}{\partial(\partial_t q_i)}. \tag{1.91}$$

For the Lagrangian (1.87) we get $\pi = m\partial_t q$. For the Lagrangian (1.89) we have two variables, $q_1 = q$ and $q_2 = p$. The corresponding conjugated momenta are $\pi(q) = p$ and $\pi(p) = 0$ (there is no momentum associated with the momentum!). Once we have the momentum we may construct the Hamiltonian from the Lagrangian using the Legendre transform:

$$H = \sum_i \pi_i \partial_t q_i - L \tag{1.92}$$

For the Lagrangian (1.87) we get

$$H = \frac{p^2}{2m} + V(q),$$

whereas for the Lagrangian (1.89) we get

$$H = \pi(q)\partial_t q + \pi(p)\partial_t p - L = p\partial_t q + 0 - p\partial_t q + H = H,$$

as of course expected.

Now consider Schrödinger's equation (1.80) and let us write it in terms of the components $\psi_n$ in some basis:

$$i\partial_t \psi_n = \sum_m H_{n,m}\psi_m, \tag{1.93}$$

where $H_{n,m} = \langle n|H|m\rangle$. We now ask the following question: can we cook up a Lagrangian and an action such that the corresponding Euler-Lagrange equations give Eq. (1.93)? The answer, of course, is yes.[2] The "variables" in this case are all components $\psi_n$. But since they are complex variables, we actually have $\psi_n$ and $\psi_n^*$ as an independent set. That is, $L = L(\psi_n, \partial_t\psi_n, \psi_n^*, \partial_t\psi_n^*)$. and the action is

$$S[\psi_n^*, \psi_n] = \int_{t_1}^{t_2} L(\psi_n, \partial_t\psi_n, \psi_n^*, \partial_t\psi_n^*)\,\mathrm{d}t. \tag{1.94}$$

The correct Lagrangian we should use is

$$L = \sum_n i\psi_n^*\partial_t\psi_n - \sum_{n,m} H_{n,m}\psi_n^*\psi_m. \tag{1.95}$$

where $\psi_n$ and $\psi_n^*$ are to be interpreted as independent variables. Please take notice of the similarity with Eq. (1.89): $\psi_n$ plays the role of $q$ and $\psi_n^*$ plays the role of $p$. To check that this works we use the Euler-Lagrange equations for the variable $\psi_n^*$:

$$\frac{\partial L}{\partial\psi_n^*} - \partial_t\left(\frac{\partial L}{\partial(\partial_t\psi_n^*)}\right) = 0.$$

The second term is zero since $\partial_t\psi_n^*$ does not appear in Eq. (1.95). The first term then gives

$$\frac{\partial L}{\partial\psi_n^*} = i\partial_t\psi_n - \sum_m H_{n,m}\psi_m = 0.$$

which is precisely Eq. (1.93). Thus, we have just cast Schrödinger's equation as a principle of least action for a weird action that depends on the quantum state $|\psi\rangle$. I will leave to you as an exercise to compute the Euler-Lagrange equation for $\psi_n$; you will simply find the complex conjugate of Eq. (1.93).

Eq. (1.95) is written in terms of the components $\psi_n$ of a certain basis. We can also write it in a basis independent way, as

$$\boxed{L = \langle\psi|(i\partial_t - H)|\psi\rangle} \tag{1.96}$$

This is what I call the Schrödinger Lagrangian. Isn't it beautiful? If this abstract version ever confuse you, simply refer back to Eq. (1.95).

Let us now ask what is the conjugated momentum associated with the variable $\psi_n$ for the Lagrangian (1.95). Using Eq. (1.91) we get,

$$\pi(\psi_n) = \frac{\partial L}{\partial(\partial_t\psi_n)} = i\psi_n^*, \qquad \pi(\psi_n^*) = 0 \tag{1.97}$$

[2] If the answer was no, I would be a completely crazy person, because I just spent more than two pages describing Lagrangian mechanics, which would have all been for nothing.

This means that $\psi_n$ and $i\psi_n^*$ are conjugated variables. As a sanity check, we can now find the Hamiltonian using the definition (1.92):

$$H = \sum_n i\psi_n^* \partial_t \psi_n - L \tag{1.98}$$

which, substituting (1.95) gives just the actual Hamiltonian.