# Contents

# Chapter 2

# Quantum states and observables

Let us start by discussing the Stern-Gerlach boxes of Chapter 1 again, now in a slightly more formal language. In order to do so, we will make use of the Dirac notation. This notation is commonly used in physics nowadays. On the negative side, this chapter will contain a lot of material, much of which may appear at first a bit alien and unusual. On the positive side, once we are done with this chapter, we have pretty much covered all important structure elements of quantum mechanics. In other words, from this chapter on, we will primarily be concerned with applications – quite a consoling state of affairs. At the risk of introducing a mild redundancy here, we will follow a didactical approach: We will first discuss the formalism at hand of our example of the above boxes, and then turn to a more general formalism.

## 2.1 Quantum states of spin-$1/2$ particles

### 2.1.1 State vectors of spin degrees of freedom

Subsequently, we will abstract from the motion of the particles for a moment and merely speak about the internal degree of freedom of the particle, its spin. We will associate the situation of the spin of the particle pointing up ("spin up") with the vector

$$|0\rangle, \tag{2.1}$$

whereas the spin of the particle pointing down ("spin down") will correspond to

$$|1\rangle. \tag{2.2}$$

This is nothing to be afraid about: These are merely vectors. The reflect whether this spin is pointing up or down, and this is encoded in the $0$ or $1$. Alternatively,

we could also have written

$$|\uparrow\rangle, \tag{2.3}$$

and

$$|\downarrow\rangle, \tag{2.4}$$

but let us stick with the numbers for our purposes.

More precisely, they are basis vectors in a complex vector space $\mathcal{H}$, as we will see in a second. As such, not only vectors $|0\rangle$ and $|1\rangle$ make sense, but in fact any *linear superposition*

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \in \mathcal{H} \simeq \mathbb{C}^2, \tag{2.5}$$

where $\alpha, \beta \in \mathbb{C}$ such that

$$|\alpha|^2 + |\beta|^2 = 1. \tag{2.6}$$

For example

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \tag{2.7}$$

is a legitimate choice. Or

$$|\psi\rangle = \sin(x)|0\rangle + \cos(x)|1\rangle \tag{2.8}$$

for $x \in [0, 2\pi)$. Similarly,

$$|\times\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + i|1\rangle\right) \tag{2.9}$$

and

$$|\odot\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - i|1\rangle\right) \tag{2.10}$$

are valid choices. The parameters $\alpha$ and $\beta$ are hence complex, and Eq. (2.6) reflects normalization of the vector,

$$\||\psi\rangle\|^2 = \langle\psi|\psi\rangle = |\alpha|^2 + |\beta|^2. \tag{2.11}$$

Superpositions we have already seen before, here with complex coefficients. They are called *state vectors* in quantum mechanics. Here, they arise in a quite subtle fashion, however: A situation of the form as in Eq. (2.7) does *not* correspond to the situation of the spin pointing either up or down. It is something very different. It is a *superposition* of the spin pointing up and down (omitting the hyphenation from now on). We will see what this means very soon.

We have said that $|0\rangle$ and $|1\rangle$ correspond to spin up and down, respectively. As we will see in more detail later, coming back to the above boxes,

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \tag{2.12}$$

can be identified with spin left and

$$|-\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right) \tag{2.13}$$

with spin right. These are all state vectors of a single spin degree of freedom. Again as a box:

**State vectors of spins:** The state vector of a spin degree of freedom of a spin-$1/2$ particle can be written as a vector

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{2.14}$$

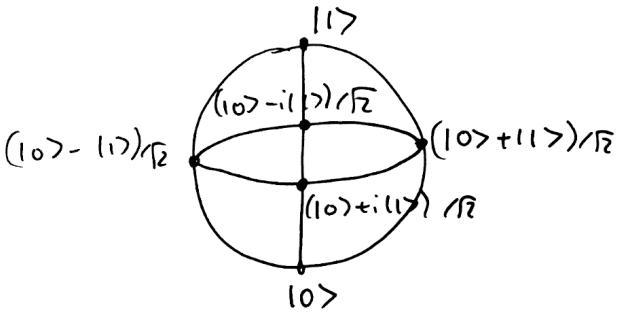
satisfying $|\alpha|^2 + |\beta|^2 = 1$.

In fact, if we can say this now without sounding too confusing: this is a mild over-paramatrization of a state in quantum theory. The global phase of a state vector and does not enter any prediction of quantum theory. That is to say, $|\psi\rangle$ and

$$e^{i\phi}|\psi\rangle,\ \phi \in \mathbb{R}, \tag{2.15}$$

are identified with the same quantum states.

As such, we immediately come to a handy and intuitive picture representing a state vector of a single spin degree of freedom: Numbers $\alpha, \beta \in \mathbb{C}$ satisfying Eq. (2.6) – up to their irrelevant global phase – can be identified with points in $\mathbb{R}^3$ on a unit sphere, so in three spatial coordinates. Spin degrees of freedom are hence like "arrows" pointing into some direction. This image is frequently employed in intuitive explanations. $|0\rangle$ and $|1\rangle$ correspond to the north and south pole of the sphere. $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle + |1\rangle/\sqrt{2}$ as well as $(|0\rangle + i|1\rangle)/\sqrt{2}$ and $(|0\rangle - i|1\rangle)/\sqrt{2}$ are on the equator of the sphere. But in fact, any point on the sphere corresponds to a legitimate state vector. This representation is referred to as the *Bloch sphere*.




Let us conclude this subsection by noting that since these vectors form a vector space, if $|\psi\rangle$ is a state vector and $|\phi\rangle$ is another state vector, then any

$$\alpha|\psi\rangle + \beta|\phi\rangle, \tag{2.16}$$

with $\alpha, \beta \in \mathbb{C}$, appropriately normalized, is again a legitimate state vector.

### 2.1.2 Pauli operators

We have seen – even if not discussed in all detail – that the internal states can be associated with vectors. We will now see that physics properties can be associated with operators $A$, so with linear maps of the form

$$|\psi\rangle \mapsto A|\psi\rangle, \tag{2.17}$$

where again $A|\psi\rangle \in \mathcal{H}$. Before we become too abstract at this point, let us consider an example: The *Pauli-$z$-matrix* $\sigma_z$ acts as

$$\sigma_z|0\rangle = |0\rangle, \tag{2.18}$$
$$\sigma_z|1\rangle = -|1\rangle. \tag{2.19}$$

We hence note that the vectors $|0\rangle$ and $|1\rangle$ are *eigenvectors* of $\sigma_z$: Up to a complex number – here $+1$ and $-1$, the respective *eigenvalues* – we obtain again the same vector if we apply $\sigma_z$ to it. The above mentioned state vectors that are on the north and south poles of the Bloch sphere hence correspond to eigenvectors of $\sigma_z$ with the respective eigenvalues. Similarly, we find that the *Pauli-$x$ matrix* $\sigma_x$ acts as

$$\sigma_x|+\rangle = |+\rangle, \tag{2.20}$$
$$\sigma_x|-\rangle = -|-\rangle. \tag{2.21}$$

Finally, the *Pauli-$y$-matrix* $\sigma_y$ has the property that

$$\sigma_y|\times\rangle = |\times\rangle, \tag{2.22}$$
$$\sigma_y|\odot\rangle = -|\odot\rangle. \tag{2.23}$$

### 2.1.3 A preliminary interpretation of the boxes

We will explain this in more detail below. The short and somewhat elliptic explanation of the above situation involving the boxes is as follows: Properties are associated with operators, in fact with Hermitian operators, see below. Such Hermitian operators are called *observables*. Pauli operators are examples of Hermitian operators. A measurement along the $z$ axis corresponds to the Pauli-$z$-matrix, and similarly for the other Pauli matrices. So the first measurement corresponds to a "measurement of the observable Pauli-$z$". After the measurement, the system will be in an eigenvector of the respective observable, the outcome of the measurement being the eigenvalue of the eigenvector. For example, we measure the spin along the $z$ axis, hence "measure the observable Pauli-$z$". If we get the value $+1$, we will obtain the state vector

$$|0\rangle = \sigma_z|0\rangle \tag{2.24}$$

after the measurement, corresponding to spin up. In case of the value $-1$, we obtain the state vector

$$|1\rangle = -\sigma_z|1\rangle \tag{2.25}$$

after the measurement. Since $|0\rangle$ is an eigenvector of $\sigma_z$, if we will repeat measuring $\sigma_z$, we will repeatedly get the outcome $+1$ and the post measurement state vector $|0\rangle$. This is why then the spin is always pointing into the same direction.

If we at some point measure along the $x$ axis, the situation is quite different. Neither $|0\rangle$ not $|1\rangle$ are eigenvectors of $\sigma_x$. We will see that this fact is essentially responsible for the two outcomes, spin left and spin right, being obtained in a probabilistic fashion. After the measurement, the state vector is an eigenstate of the observable measured, i.e.,

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) = \sigma_x \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \tag{2.26}$$

in case of one outcome and

$$|-\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right) = -\sigma_x \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right) \tag{2.27}$$

in case of the other. After this teaser, we are in the position to consider this situation a bit more systematically and more carefully.

## 2.2 States and observables

### 2.2.1 Bras and kets

In the above example, we had two basis vectors $|0\rangle$ and $|1\rangle$. Needless to say, there are situations in physics where one has a larger number of basis vectors. For example, the two levels could not only represent the spin degree of freedom, but in fact any two internal degrees of freedom. This could be the two energy levels of an atom. But having said that, there is no need for this number necessarily being two. In fact, usually, atoms have a large number of energy levels, which could be counted in one way or the other. Let us hence say that we have $d$ levels, where $d$ is any natural number. This is also called an finite-dimensional quantum system, in contrast to so-called infinite-dimensional quantum systems (the position of a particle does, for example, does not correspond to a finite-dimensional quantum system). At this point, we can surely guess how a general state vector of such a quantum system would look like. This is the most general form of a superposition.

**State vectors of finite-dimensional quantum systems:** The state vector of a $d$-dimensional quantum system can be written as a vector

$$|\psi\rangle = \sum_{j=0}^{d-1} \alpha_j |j\rangle, \tag{2.28}$$

satisfying normalization

$$\sum_{j=0}^{d-1} |\alpha_j|^2 = 1. \tag{2.29}$$

Indeed, the vectors

$$\{|0\rangle, \ldots, |d-1\rangle\} \tag{2.30}$$

are again orthonormal basis vectors, while all other state vectors are superpositions thereof. The principle that any of these vectors are allowed state vectors is sometimes also referred to as the *superposition principle*. These vectors form a vector space $\mathcal{H}$, in fact a *Hilbert space* (see Appendix).

**Hilbert space of $d$-dimensional quantum systems:** The basis vectors $\{|0\rangle, \ldots, |d-1\rangle\}$ span the Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$.

Very important will also be the scalar product.

**Scalar product:** For two state vectors

$$|\psi\rangle = \sum_{j=0}^{d-1} \alpha_j |j\rangle, \;\; |\phi\rangle = \sum_{j=0}^{d-1} \beta_j |j\rangle \tag{2.31}$$

we write their standard scalar product as

$$\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^* = \sum_{j=0}^{d-1} \alpha_j^* \beta_j. \tag{2.32}$$

Such scalar products are sometimes called "brackets", which is why vectors – in an instance of physics humor, judge for yourself – are called "kets". Dual vectors, from the dual space $\mathcal{H}^*$, are referred to as "bras",

$$\langle\psi| = \sum_{j=0}^{d-1} \beta_j \langle j|, \tag{2.33}$$

and again they are normalized if

$$\sum_{j=0}^{d-1} |\beta_j|^2 = 1. \tag{2.34}$$

As always, the scalar product induces a norm of the vector

$$\||\psi\rangle\|^2 = \langle\psi|\psi\rangle = \sum_{j=0}^{d-1} |\alpha_j|^2. \tag{2.35}$$

By construction, this is always a non-negative real number, and equal to unity in case of a normalized state vector. We have that the basis vectors are orthogonal,

$$\langle j|k\rangle = \delta_{j,k} \tag{2.36}$$

for $j, k = 0, \ldots, d-1$. They are complete, in that the span of these vectors is the entire Hilbert space, or equivalently,

$$\sum_{j=0}^{d-1} |j\rangle\langle j| = \mathbb{1}. \tag{2.37}$$

This may at this point look a bit awkward, but we will see how useful this *completeness relation* is in a minute.

Again, the global phase does not enter any prediction, so the *state* of a quantum system is strictly speaking not defined by the state vector, but by the *ray* associated with this state vector:

**Rays as quantum states:** The ray of any state vector $|\psi\rangle \in \mathcal{H}$ is defined as

$$\{\lambda|\psi\rangle : \lambda \in \mathbb{C}\}. \tag{2.38}$$

Once one has grasped this, this fact is often silently omitted, in order not to make the notation overly bombastic. This fact one usually speaks of state vectors as of "states", which is a bit imprecise in that the global phase is not defined in a state.

Summarizing this subsection, we have a $d$-dimensional complex vector space with an orthonormal basis. So we can associate

$$|\psi\rangle = \sum_{j=0}^{d-1} \alpha_j |j\rangle \tag{2.39}$$

with

$$\begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_{d-1} \end{bmatrix}, \tag{2.40}$$

and a dual vector

$$\langle\psi| = \sum_{j=0}^{d-1} \beta_j \langle j| \tag{2.41}$$

with

$$[\beta_0 \ \dots \ \beta_{d-1}]\,. \tag{2.42}$$

### 2.2.2 Observables

Observables, so "observable quantities", entities that capture a measurement prescription, are Hermitian operators. This is such an important statement that we give it a box:

**Observables:** Observables in quantum mechanics correspond to Hermitian operators,

$$A = A^\dagger. \tag{2.43}$$

The adjoint can be defined via

$$\langle\psi|(A|\phi\rangle) = (\langle\psi|A)|\phi\rangle \tag{2.44}$$

for all $|\psi\rangle, |\phi\rangle$. Such observables can again be written in matrix form: In terms of the basis $\{|0\rangle, \dots, |d-1\rangle\}$, the matrix form has the entries

$$A_{j,k} = \langle j|A|k\rangle. \tag{2.45}$$

It is clear that this matrix form uniquely characterizes the operator: We can make use of the "insertion trick"

$$A = \mathbb{1} \cdot A \cdot \mathbb{1} = \sum_{j,k} \langle j|A|k\rangle |j\rangle\langle k|. \tag{2.46}$$

In the following, we will always identify an operator with its matrix form. This is a very common identification: We will write $A$ both for the linear operator itself as well as for the matrix that reflects it given a basis. This is so common and natural that any other choice would unnecessarily make the notation very cumbersome.

As an exercise, we formulate the matrix form of the Pauli operators. The first one is very much obvious: We can write

$$\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{2.47}$$

The vectors $|0\rangle$ and $|1\rangle$ are already eigenvectors of $\sigma_z$, so we should not be surprised to see that the matrix form is diagonal. Similarly, we find the matrix

form of the other two Pauli operators. We collect them in a box:

**Pauli matrices:**

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},\ \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},\ \mathbb{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.48}$$

We have included in this list the identity $\mathbb{1}$ to the matrix form of the Pauli operators, the *Pauli matrices*, primarily for reasons of convenience. We note a number of things: First, a rather formal observation: The Pauli matrices form a basis of all Hermitian $2 \times 2$ matrices: In other words, every Hermitian $2 \times 2$ matrix $A$ can be written as

$$A = \alpha\sigma_x + \beta\sigma_y + \gamma\sigma_z + \delta\mathbb{1}, \tag{2.49}$$

with real $\alpha, \beta, \gamma, \delta$.

A profound insight is that these matrices do not commute: It is by no means the same whether we compute

$$\sigma_x\sigma_z = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{2.50}$$

or

$$\sigma_z\sigma_x = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.51}$$

The order of multiplication matters! To make this difference manifest, one considers the *commutator*

$$[\sigma_x, \sigma_z] = \sigma_x\sigma_z - \sigma_z\sigma_x, \tag{2.52}$$

which is in general different from zero.

**Commutator:** For operators $A$ and $B$, the commutator is defined as

$$[A, B] = AB - BA. \tag{2.53}$$

Properties in the quantum world correspond to operators that do not commute. This basic feature of the formalism is at the root of the observation that orders of measurements matter. It is also the key to the understanding that it does not make sense in quantum theory to think of two different quantities corresponding to non-commuting observables to "take specific values". They do not.

This does not mean, of course, that no two observables necessarily commute. This definition we emphasize here simply because this notion is so often used in the literature.

**Compatible observables:** Two observables $A$ and $B$ are compatible, if

$$[A, B] = 0. \tag{2.54}$$

### 2.2.3 Eigenvectors and eigenvalues

Indeed, the fact that observables are Hermitian means that their eigenvalues are real.

**Eigenvalue decomposition of observables:** Every (finite-dimensional) observable $A$ can be written as

$$A = \sum_{j=0}^{d-1} \lambda_j |\phi_j\rangle\langle\phi_j|, \tag{2.55}$$

where the eigenvalues satisfy $\lambda_0, \ldots, \lambda_{d-1} \in \mathbb{R}$ and the eigenvectors $\{|\phi_j\rangle\}$ can be chosen to form an orthonormal basis of $\mathcal{H}$.

This does not mean, of course, that all of the $\{\lambda_j\}$ are necessarily different. If two or more eigenvalues are identical, they are called degenerate. The respective eigenvalues then span the eigenspace, which is no longer one-dimensional. Eigenvalues and eigenvectors play a very important role in quantum mechanics: The former essentially as "measurement outcomes" and the latter as "post-measurement state vectors". Since eigenvalue decompositions are so important, we state at this point an equivalent form of the eigenvalue decomposition of observables:

**Diagonalization of observables:** Every (finite-dimensional) observable $A$ can be written as

$$A = UDU^\dagger, \tag{2.56}$$

where $D = \mathrm{diag}(\lambda_0, \ldots, \lambda_{d-1})$, $\lambda_0, \ldots, \lambda_{d-1} \in \mathbb{R}$, and $U$ is unitary, so satisfied

$$UU^\dagger = U^\dagger U = \mathbb{1}. \tag{2.57}$$

Unitary operators are those that preserve scalar products, and correspond to basis changes of orthonormal basis. The above statement is hence a manifestation of the observation that in their eigenbasis, observables are diagonal. The

above notation is also common for

$$D = \text{diag}(\lambda_0, \ldots, \lambda_{d-1}) = \sum_{j=0}^{d-1} \lambda_j |j\rangle\langle j|. \tag{2.58}$$

Hermitian observables can hence always be unitarily diagonalized. This is even true for slightly more general operators: The same statement in Eq. (2.56) is true for so-called normal operators $A$ which satisfy $AA^\dagger = A^\dagger A$. This is a helpful insight, in particular when one thinks of the diagonalization of unitary operators $U$ (which are in general not Hermitian): This can be written as

$$U = \sum_{j=0}^{d-1} e^{i\phi_j} |j\rangle\langle j|, \tag{2.59}$$

with all $\phi_j \in \mathbb{R}$. The eigenvalues are hence distributed on the unit circle in the complex plane.

We also emphasize an important insight: Observables can be simultaneously (unitarily) diagonalized if and only if they commute. We end this subsection with a note again on the (three, different from $\mathbb{1}$) Pauli matrices: Their determinant and their trace are given by

$$\det(\sigma_i) = -1, \; \text{tr}(\sigma_i) = 0 \tag{2.60}$$

for $i = x, y, z$, and hence all eigenvalues are $\pm 1$.

### 2.2.4 Position representation and wave functions

Before we finally come to the role of measurement and can carefully and properly explain our above setting involving boxes, let us turn to another important aspect of the above situation. We have neglected the spatial degree of freedom and have looked at the spin degree of freedom alone. Let us now do the opposite and let us look at a quantum mechanical description of a spatial degree of freedom alone.

Most quantum mechanics books start here. This may be taken as an intuitive approach, as ideas of particles flying around in free space is very intuitive. Also, the idea of a wave function, allowing for an interpretation of a density, is probably already quite familiar. Unfortunately, the treatment of position and momentum is somewhat overburdened with mathematical technicalities, if one wants to do things rigorously. We take a pragmatic approach here, and will try to steer away from any fine print, without saying anything wrong at any instance in time. Let us start with the operators reflecting position and momentum:

**Position and momentum:** Position and momentum are reflected by the Hermitian position and momentum operators $X$ and $P$.

Now, these operators no longer have discrete spectra, so no discrete eigenvalues, as we now have $\dim(\mathcal{H}) = \infty$. We can at this point already grasp why this is the case: eigenvalues correspond to measurement outcomes. In case of position, a particle can take a continuum of different positions, in case of a one-dimensional problem, this is the entire real line. And in fact, both the position and the momentum operator have $\mathbb{R}$ as their spectrum.

Neither the position nor the momentum operator have eigenvectors (although many books spell them out nevertheless, we will come back to that). It still makes sense for $x \in \mathbb{R}$ to write

$$\langle x|\psi\rangle = \psi(x), \tag{2.61}$$

even if at this point we take the right hand side to be the definition for the left hand side of the equality sign (we do not insist the dual vectors $\langle x|$ to be contained in any dual Hilbert space, but we start worrying about that detail in the appendix). This representation in terms of wave functions is usually called the *position representation*.

**Wave functions:** Single spatial degrees of freedom are described by wave functions

$$\psi : \mathbb{R} \to \mathbb{C}. \tag{2.62}$$

They satisfy

$$\int_{-\infty}^{\infty} |\psi(x)|^2 = 1. \tag{2.63}$$

$|\psi(x)|^2$ can be interpreted as the probability density for finding the particle at position $x$.

Because we ask for $\psi$ being square integrable, the Hilbert space is $\mathcal{H} = L^2(\mathbb{R})$ of square integrable functions over the reals. So indeed,

$$\rho(x) = |\psi(x)|^2 \tag{2.64}$$

for $x \in \mathbb{R}$ has a probability interpretation. Yet, the wave function is "a lot more than a mere probability distribution", as it also contains phase information, as it is a complex-valued function. In other words, the wave function does not reflect the situation that "the particle is somewhere, we merely do not know where it is". This statement – even if one can sometimes read such statements in the literature – is wrong. What is true, however, is that the probability density of finding particles in particular places can be determined from the wave function.

Again, state vectors are elements of Hilbert (and hence vector) spaces: That is to say that if $|\psi_1\rangle$ is a legitimate state vector and $|\psi_2\rangle$ as well, then any linear combination of them, appropriately normalized, is again a state vector. Stated in terms of the position representation, if $\psi_1 : \mathbb{R} \to \mathbb{C}$ and $\psi_2 : \mathbb{R} \to \mathbb{C}$ are wave

functions, so is $\psi = \psi_1 + \psi_2$ (up to normalization), so the function with values

$$\psi(x) = \frac{1}{\sqrt{N}} \left(\psi_1(x) + \psi_2(x)\right), \tag{2.65}$$

where $N$ reflects normalization,

$$N = \int_{-\infty}^{\infty} dx \left|\psi_1(x) + \psi_2(x)\right|^2. \tag{2.66}$$

This is by no means a detail: The probability density $\rho(x)$ of finding a particle at $x$ is now given by

$$\begin{aligned} \rho(x) &= \frac{1}{\sqrt{N}} |\psi_1(x) + \psi_2(x)|^2 \\ &= \frac{1}{\sqrt{N}} \left(|\psi_1(x)|^2 + |\phi_2(x)|^2 + \psi_1^*(x)\psi_2(x) + \psi_1(x)\psi_2^*(x)\right). \end{aligned} \tag{2.67}$$

This is *different* from

$$\frac{1}{\sqrt{N}} \left(|\psi_1(x)|^2 + |\psi_2(x)|^2\right), \tag{2.68}$$

in that new terms arise reflecting the superposition. This is exactly the feature that is seen in the familiar double slit experiment: One does not merely see the joint pattern of each of the preparations in a superposition: But in fact, new features emerge. Note also that this is not a statistical interference because of many particles coming together or so. A single particle, suitable prepared, will show the interference pattern, reflecting the fact that the complex phases matter.

Again, one can take an orthonormal basis of the Hilbert space $\mathcal{H}$ of the spatial degree of freedom of a particle, but this is no longer finite-dimensional, but infinite-dimensional. One has

$$\mathcal{H} = \overline{\text{span}}\{|0\rangle, |1\rangle, \dots\}, \tag{2.69}$$

where the bar marks that the Hilbert space is given by the closure of the span of the basis vectors.

### 2.2.5 Momentum representation and Fourier transforms

Position and momentum do not commute. This is often – and famously – written as

$$[X, P] = i\hbar. \tag{2.70}$$

Some caution is needed here, as neither the position nor the momentum operators are bounded operators. Again, we will offer a fix for that in the appendix.

Similarly to the position representation, we can introduce the *momentum representation*: Similarly as above, we consider

$$\langle p|\psi\rangle = \tilde{\psi}(p), \tag{2.71}$$

where $\psi : \mathbb{R} \to \mathbb{C}$ is the Fourier transform of $\tilde{\psi} : \mathbb{R} \to \mathbb{C}$,

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{\psi}(k) e^{ikx} dk. \tag{2.72}$$

### 2.2.6 Combining quantum systems

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

Composition of degrees of freedom is incorporated by the tensor products in quantum mechanics. Let us assume that we have one degree of freedom associated with a $d_1$-dimensional Hilbert space

$$\mathcal{H}_1 = \mathrm{span}\{|0\rangle, \ldots, |d_1 - 1\rangle\}. \tag{2.73}$$

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

$$\mathcal{H}_2 = \mathrm{span}\{|0\rangle, \ldots, |d_2 - 1\rangle\}. \tag{2.74}$$

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

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{2.75}$$

It is spanned by the orthonormal basis vectors

$$\{|j\rangle \otimes |k\rangle : j = 0, \ldots, d_1 - 1; k = 0, \ldots, d_2 - 1\}. \tag{2.76}$$

Such basis elements of tensor products are sometimes also written as

$$\{|j, k\rangle : j = 0, \ldots, d_1 - 1; k = 0, \ldots, d_2 - 1\}. \tag{2.77}$$

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

$$|\psi_1\rangle = \sum_{j=0}^{d_1} \alpha_j |j\rangle \tag{2.78}$$

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

$$|\psi_2\rangle = \sum_{j=0}^{d_2} \beta_j |j\rangle, \tag{2.79}$$

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

$$|\psi\rangle = \sum_{j=0}^{d_1} \sum_{k=0}^{d_2} \gamma_{j,k} |j\rangle \otimes |k\rangle, \tag{2.80}$$

as a linear combination of all new basis vectors, with all $\gamma_{j,k} \in \mathbb{C}$. If you think at this point that it may be confusing that such general state vectors contain ones that are no longer a product between the respective Hilbert spaces: Indeed, it is, and we will come to the profound implications of this later. Again:

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

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{2.81}$$

Similarly, an arbitrary linear operator can be decomposed as

$$O = \sum_{j,k} c_{j,k} A_j \otimes B_k, \tag{2.82}$$

with operators $\{A_j\}$ and $\{B_k\}$ on $\mathcal{H}_1$ and $\mathcal{B}_1$, respectively.

The same is true, needless to say, for more elaborate composite quantum systems having many parts. For example, the Hilbert space of $n$ particles of the same character (and associated with $\mathcal{H}$ each) is given by

$$\mathcal{H} \otimes \cdots \otimes \mathcal{H} = \mathcal{H}^{\otimes n}. \tag{2.83}$$

This is not such an alien situation in quantum physics: We should not forget that most things we can see are quantum systems with more than $10^{23}$ individual parts.

The parts of composite quantum systems also do not have to be the same. In fact, on the first page of this script we have already encountered a composite quantum system, even if at this point we had abstracted from this fact. The particle has a position and a spin degree of freedom. So the joint Hilbert space capturing this situation is given by

$$\mathcal{H} = L^2(\mathbb{R}) \otimes \mathbb{C}^2. \tag{2.84}$$

The Hilbert space of the three coordinates of the spatial degree of freedom of a single particle is given by

$$\mathcal{H} = L^2(\mathbb{R}) \otimes L^2(\mathbb{R}) \otimes L^2(\mathbb{R}). \tag{2.85}$$

## 2.3 Measurement

After all this, we are finally in the position to carefully and precisely describe the boxes that we have encountered above. For doing so, we need to precisely understand what measurement does to a quantum system.

### 2.3.1 Measurement postulate

Let us start by precisely stating how measurement is captured in quantum theory, and then discuss what this means in great detail. The subsequent type of measurement is called "von-Neumann measurement" (to emphasize that later on, we will have a look at a more general framework to capture measurement, which is however not fundamentally more general than the measurement postulate that we are going to see now).

**Measurement postulate (for von-Neumann measurements):** We consider a measurement of the observable $A$ for a quantum system prepared in a state vector $|\psi\rangle \in \mathbb{C}^d$. Let us assume that none of the eigenvalues $\lambda_0, \ldots, \lambda_{d-1}$ of

$$A = \sum_{j=0}^{d-1} \lambda_j |\phi_j\rangle\langle\phi_j| \tag{2.86}$$

are degenerate. Then the probability $p_j$ of obtaining the outcome $\lambda_j$ is given by

$$p_j = |\langle\psi|\phi_j\rangle|^2, \tag{2.87}$$

the state vector immediately after the measurement is given by

$$|\psi_j\rangle = \frac{1}{\sqrt{p_j}} \langle\phi_j|\psi\rangle |\phi_j\rangle. \tag{2.88}$$

In other words, the measurement values are the eigenvalues, and a measurement "collapses" the state vector into one of the eigenvectors. The situation of encountering degenerate eigenvalues merely requires a mild modification: In case of a degenerate observable $A$, with eigenvalues $\lambda_0, \ldots, \lambda_{d-1}$ and different eigenvalues $\mu_0, \ldots, \mu_{D-1}$, we can write

$$A = \sum_{k=0}^{D-1} \mu_k \pi_k \tag{2.89}$$

where

$$\pi_k = \sum_{j, \lambda_j = \mu_k} |\phi_j\rangle\langle\phi_j| \tag{2.90}$$

is for each $k$ the orthogonal projection onto the subspace spanned by the eigenvalues that take the value $\mu_k$. As these are projections, we have

$$\pi_k = \pi_k^2, \tag{2.91}$$
$$\pi_k = \pi_k^\dagger \tag{2.92}$$

for all $k$. Then the probability of obtaining the outcome $\mu_k$, $k = 0, \ldots, D-1$, is given by

$$p_k = \langle\psi|\pi_k|\psi\rangle, \tag{2.93}$$

the state vector immediately after the measurement is given by

$$|\psi_k\rangle = \frac{1}{\sqrt{p_k}}\pi_k|\psi\rangle. \tag{2.94}$$

Clearly, for non-degenerate observables, this prescription simply becomes the previous one.

### 2.3.2 Expectation values

Needless to say, a single measurement will not provide us with complete information about the probabilities $\{p_j\}$ – similarly to a single coin being tossed will not say whether the coin is fair. The expectation value is of key importance in quantum mechanics: is the expected value of the averaged outcomes of a single type of measurement.

**Expectation values:** The expectation value of an observable $A$ for a state vector $|\psi\rangle$ is given by

$$\langle A\rangle = \langle\psi|A|\psi\rangle. \tag{2.95}$$

Because observables are Hermitian, expectation values are always real.

For example, if we measure a Pauli operator $\sigma_z$ on a system initially prepared in $|0\rangle$. Then we will obtain the expectation value

$$\langle 0|\sigma_z|0\rangle = 1. \tag{2.96}$$

In case of $|1\rangle$, we get

$$\langle 1|\sigma_z|1\rangle = -1. \tag{2.97}$$

And in case of $|+\rangle$ as defined in Eq. (2.7), we have

$$\langle +|\sigma_z|+\rangle = 0. \tag{2.98}$$

Again, a single run of the experiment will not provide us the expectation value, but the averaged outcomes of many runs will asymptotically converge to this value.

### 2.3.3 A proper explanation of the boxes

We are now in the position to properly explain our initial situation involving boxes – with the single exception of the initial preparation, to which we will come later. The first measurement of the Pauli operator $\sigma_z$ can have the outcomes $\lambda_0, \lambda_1 = \pm 1$. After the measurement, the wave function is projected onto $|0\rangle$ or $|1\rangle$, dependent on the measurement outcome. This is why any subsequent measurement of $\sigma_z$ will provide the same outcome: One the state vector is collapsed to $|0\rangle$, say, the probability of obtaining the outcome $1$ is again

$$p_0 = 1. \tag{2.99}$$

A fancy way of saying this is that $\sigma_z$ and $\sigma_z$ are compatible, and we can repeat the measurement and arbitrary number of times, and will get the same value over and over again.

This is different in case we measure $\sigma_x$ in between. Say, we have obtained the outcome $\lambda_0 = 1$ and the post-measurement state $|0\rangle$. Then the probability of getting the post-measurement-state $|+\rangle$ is given by

$$|\langle +|0\rangle|^2 = \frac{1}{2}, \tag{2.100}$$

and in the same way

$$|\langle -|0\rangle|^2 = \frac{1}{2}. \tag{2.101}$$

So the probability of getting each outcome is precisely $1/2$. Then, if we obtain $|+\rangle$ in the measurement of the observable $\sigma_x$, then the probability of getting $|0\rangle$ in a subsequent $\sigma_z$ measurement is again $1/2$, for the same reasons – only with the roles of the two vectors reversed. Having understood the measurement postulate, the above situation is very clear.

This notion of a measurement of $\sigma_x$ "disturbing" a measurement of $\sigma_z$ is a manifestation of the observables $\sigma_x$ and $\sigma_z$ not commuting: They are incompatible. Compatible observables can be repeatedly measured without altering the outcome, while this is not so for incompatible observables. The order of measurement also matters, and it does simply not make sense to ask whether the spin is "truly pointing up or down" in case one has just performed a measurement along the $x$ axis. Such properties are simply not defined within quantum theory.

### 2.3.4 Three readings of the Heisenberg uncertainty relation

This notion of outcomes of measurements being "uncertain" can be made more precise in terms of the so-called uncertainty relation. There are few statements in quantum theory, yet, that are so often misunderstood as the uncertainty relation. Recently, a lawyer came to me and asked, well, is the uncertainty in adjudication not just a manifestation of the Heisenberg principle, that "nothing is precisely defined"? Well, actually, no.

But except from the principle being overly referred to in urban slang, even within the formalism there often is a kind of confusion coming along with this principle. I was slightly confused when I first heard of it in my first lecture, and slightly irritated by the fact that nobody seemed to have any desire to fix the precise prefactor on the right hand side of the inequality. Later I understood that this is because there have been different readings used of the principle. We will mention three readings of Heisenberg's uncertainty principle, even if we discuss in detail only one.

The central object here is the mean square deviation of an observable from its expectation value, for a system initially prepared in $|\psi\rangle$,

$$(\Delta A)^2 = \langle (A - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2. \tag{2.102}$$

We now consider this for two observables $A$ and $B$. For simplicity of notation, let us define

$$\begin{aligned} C &= A - \langle A \rangle, & (2.103)\\ D &= B - \langle B \rangle. & (2.104) \end{aligned}$$

Even if $C$ and $D$ are clearly Hermitian, the product $CD$ is usually no longer Hermitian, and the expectation value can be complex. Let us write

$$\begin{aligned} \langle CD \rangle &= \langle \psi | CD | \psi \rangle = z = x + iy, & (2.105)\\ \langle DC \rangle &= \langle \psi | (CD)^\dagger | \psi \rangle = z^* = x - iy, & (2.106) \end{aligned}$$

with $x, y \in \mathbb{R}$. That is to say,

$$\langle [C, D] \rangle = \langle \psi | [C, D] | \psi \rangle = z - z^* = 2iy = 2i \, \mathrm{im}(\langle CD \rangle). \tag{2.107}$$

Taking the absolute value, we obtain

$$|\langle [C, D] \rangle|^2 \leq 4 |\langle \psi | CD | \psi \rangle|^2. \tag{2.108}$$

Of course, we have for the commutator

$$[C, D] = [A, B], \tag{2.109}$$

as numbers always commute with each other. We now make use of the Cauchy-Schwarz inequality – a useful inequality from linear algebra. This leads to

$$\begin{aligned} |\langle \psi | CD | \psi \rangle|^2 &= |(\langle \psi | C) \cdot (D | \psi \rangle)|^2 \\ &\leq \| C | \psi \rangle \|^2 \cdot \| D | \psi \rangle \|^2 \\ &= \langle \psi | C^2 | \psi \rangle \cdot \langle \psi | D^2 | \psi \rangle. \end{aligned} \tag{2.110}$$

We can now collect the terms and state the uncertainty relation:

**Heisenberg's uncertainty relation:** Two observables $A$ and $B$ of a quantum system prepared in $|\psi\rangle$ satisfy the uncertainty relation

$$\Delta A \cdot \Delta B \geq \frac{1}{2}|\langle\psi|[A,B]|\psi\rangle|. \tag{2.111}$$

That is to say, the product of the two uncertainties cannot take an arbitrarily small value – unless the two observables $A$ and $B$ commute. Otherwise, the product of the two uncertainties of $A$ and $B$ cannot be arbitrarily small. This is a remarkable observation.

When applied to position and momentum, this principle reads

$$\Delta X \cdot \Delta P \geq \frac{\hbar}{2}, \tag{2.112}$$

which means that wave functions cannot be arbitrarily narrow both in position and momentum.

This means that if we prepare many quantum systems and first estimate $\Delta A$ many times, such that we know it later with high statistical significance, and then estimate $\Delta B$: Then the product of the two uncertainties will always be larger than the above given value. Quantum mechanics simply does not allow for any smaller uncertainties. So in a way, one cannot "know the values of two non-commuting observables at once". Precisely, what is meant, however, is that one independently prepares the systems and measures *either* $A$ or $B$, and then analyses the data. This is the first reading of the Heisenberg uncertainty relation.

This is also the derivation that most books on quantum mechanics offer. Interestingly, the explanation given is quite often incompatible with the above derivation. This is the notion of a measurement of one observable $A$ makes the outcome of another non-commuting observable $B$ less certain. This is the second reading of the uncertainty relation. We hence measure on the *same* system first $A$, then $B$. If we "know the value of $A$ precisely, then the measurement of $B$ will be a lot disturbed". This is the reading of the famous Heisenberg microscope: Heisenberg discussed how a measurement of $X$ disturbs later later measurements of the non-commuting observable $P$. This is also true, but this is not quite what we have derived above (which made use of the independent and identical distribution (i.i.d.) of the initial preparation). One can also derive uncertainties for the Heisenberg uncertainty principle in this second reading, but the prefactor on the right hand side will be slightly different.

A third reading which we only briefly mention is the one where tries to *jointly* measure two observables in a single generalized measurement. We have to delay this discussion as we are at this point not quite clear about what a generalized measurement is. Just for the records: Here one aims at obtaining as much information as possible about $A$ and $B$ in a single run of a measurement,

which is then repeated many times. This again leads to an uncertainty relation, but yet again with a different prefactor.

All these readings have the narrative in common, however, that if the mean square deviation of one observable is small, it cannot be small at the same time for a non-commuting observable. This is again one of the profound consequences of non-commuting operators in quantum theory.