**Advanced quantum mechanics (20104301)**
Lecturer: Jens Eisert
Chapter 1: Elements of quantum theory

# Chapter 1

# Elements of quantum theory

Every physical theory is supposed to make predictions on future measurement outcomes when performing experiments with a well-defined physical system that is initially prepared in the same way. The predictions of quantum mechanics are of a statistical nature: The theory is utterly silent about specific measurement outcomes. It will rather provide probabilities for obtaining certain outcomes. Conversely, to obtain evidence into the correctness of a prediction, one needs to perform many experiments under identical conditions. Then, by investigating relative frequencies of measurement outcomes, one can estimate probabilities. At the heart of the formalism are notions of expectation values. This is no shortcoming of the theory: This intrinsic randomness is actually a deep structure element of quantum mechanics that is there to stay: Bell's theorem shows that there cannot be an underlying classical statistical picture that can be held responsible to explain the randomness of quantum mechanics.

That is to say, we have to ask ourselves how to capture *states* – the collection of information summarizing all information required to make future predictions – how *observables* – the quantities that can be measured. Also, we will think about how systems evolve in time.

## 1.1 Quantum states and observables

### 1.1.1 Pure quantum states

We start by considering quantum systems that are not composite, but that consist of simple physical systems. Quantum systems are associated with a *Hilbert space*. A complex Hilbert space is a vector space equipped with a scalar product that is complete with respect to the norm induced by the its scalar product. The simplest conceivable Hilbert space is that of a single spin. It is spanned by the two vectors $|0\rangle$ and $|1\rangle$ or $|\uparrow\rangle$ and $|\downarrow\rangle$: That is to say, the spin points up or down.

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

This could be a superposition

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1.1}$$

with complex $\alpha, \beta$, normalized as

$$|\alpha|^2 + |\beta|^2 = 1. \tag{1.2}$$

This example already shows that spins cannot only be pointing up or down in quantum mechanics. They can be in an arbitrary superposition of pointing up or down. The associated Hilbert space is simply $\mathcal{H} = \mathbb{C}^2$.

The scalar product between two state vectors is written as $\langle\psi|\phi\rangle$. Normalization means that the standard vector norm takes a unit value, which in turn is equivalent with

$$\langle\psi|\psi\rangle = 1. \tag{1.3}$$

The vector $\langle\psi| \in \mathcal{H}^*$ is a dual vector. Jokingly referring to the term bracket, one also calls dual vectors "bras" and vectors "kets". Matrix elements of operators $A$ take the form $\langle\psi|A|\phi\rangle$.

For every Hilbert space of a $d$-dimensional quantum system, so a system with $d$ "levels", one can pick a basis

$$\mathcal{B} = \{|0\rangle, \ldots, |d-1\rangle\}. \tag{1.4}$$

In this basis, every state vector can be expressed as

$$|\psi\rangle = \sum_{j=0}^{d-1} c_j |j\rangle. \tag{1.5}$$

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

$$\langle j|k\rangle = \delta_{j,k}, \tag{1.6}$$

$$\sum_{j=0}^{d-1} |j\rangle\langle j| = \mathbb{1}. \tag{1.7}$$

All this applies to so-called finite-dimensional quantum systems, where $d$ is an integer. The Hilbert space of a *particle in the position representation* is $\mathcal{H} = L^2(\mathbb{R})$: Pure states of particles without spin in one spatial dimension are captured by by *wave function*. State vectors $|\psi\rangle$ hence then belong to complex wave functions $\psi$ that are normalized as

$$\int dx \psi^*(x)\psi(x) = 1. \tag{1.8}$$

Their scalar product is then

$$\int dx \phi^*(x)\psi(x) = \langle\phi|\psi\rangle. \tag{1.9}$$

A good chunk of most elementary quantum mechanics lectures is concerned with the dynamics of such wave functions, say, the scattering induced by a box potential. As we know,

$$p(x) = |\psi(x)|^2 \tag{1.10}$$

is the probability density of finding a particle at the position $x$ in a position measurement. Here, this refers to a single spatial dimension, but we will come to more than one spatial dimension in a second. In infinite-dimensional Hilbert spaces, such as the one reflecting our situation, it is sufficient to have a separable Hilbert space. For the situation at hand, the eigenfunctions of the *harmonic oscillator*

$$\mathcal{B} = \{|0\rangle, |1\rangle, \dots\} \tag{1.11}$$

provide such a countable (but infinite) basis.

### 1.1.2 Observables

Quantities that can be measured as called, unsurprisingly, *observables*. The are associated with Hermitian operators $A$, meaning that

$$A = A^\dagger. \tag{1.12}$$

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

**Observables:** Observable are Hermitian operations in a Hilbert space. Expectation values of such observables for systems prepared in pure states are given as

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

As mentioned before, such expectation values make predictions about relative frequencies in experiments. Observables can always be diagonalized. After all, unitary operators are precisely those that map one orthonormal basis onto another one.

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

$$A = UDU^\dagger. \tag{1.14}$$

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

can, needless to say, be expressed in this basis, as

$$A = \sum_{j,k=0}^{d-1} \langle j|A|k\rangle |j\rangle\langle k|. \tag{1.15}$$

Their *trace* is given by

$$\mathrm{tr}[A] = \sum_{j=0}^{d-1} \langle j|A|j\rangle. \tag{1.16}$$

The trace is independent of the choice of the basis, as a moment of thought reveals. Operators in infinite-dimensional Hilbert spaces do not need to have a trace. One can still define trace-class operators for which the sequence of traces of nets converges.

In what follows, we will not distinguish between operators and their matrix representation, as is common in the literature. A kind of observable that takes a key role for good reasons are the *Pauli operators*. They are defined as

$$X = |0\rangle\langle 1| + |1\rangle\langle 0| = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{1.17}$$

$$Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \tag{1.18}$$

$$Z = |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{1.19}$$

The unit operator

$$\mathbb{1} = |0\rangle\langle 0| + |1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{1.20}$$

is commonly included as a Pauli matrix. It is obvious how to compute their expectation values. The expectation value of $Z$ of a system prepared in $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is simply

$$\langle\psi|Z|\psi\rangle = |\alpha|^2 - |\beta|^2. \tag{1.21}$$

Stern-Gerlach type experiments can be described like that.

### 1.1.3 Mixed quantum states

Let us imagine we have a single spin, associated with a Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$. We now throw a coin. In case of heads, we prepare the spin in $|0\rangle$, in case of tails, we prepare it in $|1\rangle$. That is to say, with the classical probability $1/2$ we have $|0\rangle$, and with classical probability $1/2$ we get $|1\rangle$. How do we capture this situation? Can we describe the system by a state vector

$$|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}? \tag{1.22}$$

Not quite. This is easy to see: In case of a $\sigma_x$ measurement, we would always get the same outcome. But this is different from the situation we encounter here. In fact, when we make a measurement of $\sigma_x$, we would get both outcomes with equal probability. Or

$$|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}? \tag{1.23}$$

Again, this will not work, for the same reason. In fact, no state vector is associated with such a situation, and for that, we need to generalize our concept of a quantum state slightly: to density operators. This is, however, the most general quantum state in standard quantum mechanics, and we will not have to generalize it any further.

In fact, the above situation is an instance of the situation where we prepare with probability $p_j$, $j = 1, \ldots, n$, a system in a state vector $|\psi_i\rangle$. Since we encounter a probability distribution, we have

$$\sum_{j=1}^{n} p_j = 1. \tag{1.24}$$

Such a situation is sometimes referred to as a mixed ensemble. How do we incorporate that? To start off from what we know, let us express first pure states corresponding to state vectors as a density operator.

**Density operator of a pure state:** A pure state associated with a state vector $|\psi\rangle \in \mathcal{H}$ from some Hilbert space $\mathcal{H}$ is given by the density operator

$$\rho = |\psi\rangle\langle\psi|. \tag{1.25}$$

We immediately find some properties of such an operator: We obviously have that

$$\rho = \rho^\dagger. \tag{1.26}$$

Then,

$$\mathrm{tr}(\rho) = 1. \tag{1.27}$$

Finally, we have that

$$\rho \geq 0, \tag{1.28}$$

which means that all of its eigenvalues are non-negative, which is clearly the case, as all the eigenvalues are given by $0$ or $1$, clearly non-negative numbers. We also have the property that

$$\mathrm{tr}(\rho^2) = \mathrm{tr}(|\psi\rangle\langle\psi|\psi\rangle\langle\psi|) = \mathrm{tr}(|\psi\rangle\langle\psi|) = \mathrm{tr}(\rho) = 1. \tag{1.29}$$

How to we compute expectation values from such a density operator? Well, we know that for an observable

$$\langle A\rangle = \langle\psi|A|\psi\rangle. \tag{1.30}$$

This we can equally well write as

$$\langle A\rangle = \langle\psi|A|\psi\rangle = \mathrm{tr}(A|\psi\rangle\langle\psi|) = \mathrm{tr}(A\rho). \tag{1.31}$$

We have hence made use of one of the above rules, and have written expectation values as a trace of the observable, multiplied with the density operator. A general density operator is just extended by linearity from this definition.

**Density operator of a mixed ensemble:** Consider the situation of preparing $|\psi_j\rangle$, $j = 1, \ldots, n$ with probability $p_j$. This is associated with a density operator

$$\rho = \sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j|. \tag{1.32}$$

Then, how do we compute expectation values from that? We have for observables $A$ that

$$\langle A\rangle = \sum_{j=1}^{n} p_j \langle\psi_j|A|\psi_j\rangle, \tag{1.33}$$

from the very definition of a mixed ensemble. This we can, however, also write as

$$\begin{aligned} \langle A\rangle &= \sum_{j=1}^{n} p_j \mathrm{tr}(A|\psi_j\rangle\langle\psi_j|) \\ &= \mathrm{tr}\left(A\sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j|\right) \\ &= \mathrm{tr}(A\rho). \end{aligned} \tag{1.34}$$

So again, expectation values are just computable as the trace of the density operator multiplied with the observable. We now once more investigate properties of such a density operator: We find that again,

$$\rho = \rho^\dagger. \tag{1.35}$$

In the same fashion as before, since now

$$\sum_{j=1}^{n} p_j = 1, \tag{1.36}$$

we also have that

$$\mathrm{tr}(\rho) = 1. \tag{1.37}$$

Finally, we have that

$$\rho \geq 0, \tag{1.38}$$

since all of the probabilities are positive, and a sum of positive operators is positive. These are exactly the same properties as above, except from one that is now missing. We no longer have $\mathrm{tr}(\rho^2) = 1$. In fact, this property is replaced by

$$\begin{aligned} \mathrm{tr}(\rho^2) &= \mathrm{tr}\left(\sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j|\right)\left(\sum_{k=1}^{n} p_k |\psi_k\rangle\langle\psi_k|\right) \\ &= \sum_{j,k=1}^{n} p_j p_k \mathrm{tr}(|\psi_j\rangle\langle\psi_j|\psi_k\rangle\langle\psi_k|) \leq 1 \end{aligned}$$

where we have bounded the scalar products between two arbitrary state vectors. We have now arrived at the most general concept of a state in (standard) quantum mechanics. This is surely worth a box:

**Density operators:** General states of quantum systems with Hilbert space $\mathcal{H}$ are given by density operators $\rho$. Their properties are

$$\begin{aligned} \rho &= \rho^\dagger \text{ (Hermicity)}, && (1.39)\\ \rho &\geq 0 \text{ (Positivity)}, && (1.40)\\ \mathrm{tr}(\rho) &= 1 \text{ (Normalization)}. && (1.41) \end{aligned}$$

*Pure states* are those density operators for which

$$\mathrm{tr}(\rho) = 1, \tag{1.42}$$

those can be represented by state vectors $|\psi\rangle \in \mathcal{H}$ as

$$\rho = |\psi\rangle\langle\psi|. \tag{1.43}$$

Otherwise, if $\mathrm{tr}(\rho^2) < 1$, the state is called *mixed*. For observables, expectation values are computed as $\langle A\rangle = \mathrm{tr}(A\rho)$.

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

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|. \tag{1.44}$$

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

$$\rho = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{1.45}$$

We have that

$$\mathrm{tr}(\rho^2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1. \tag{1.46}$$

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

$$\rho = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{1.47}$$

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

$$\rho = \begin{bmatrix} p_0 & 0 \\ 0 & p_1 \end{bmatrix}. \tag{1.48}$$

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

$$\rho = \frac{1}{2}|+\rangle\langle+| + \frac{1}{2}|-\rangle\langle-|. \tag{1.49}$$

This is

$$\begin{aligned} \rho &= \frac{1}{4}\left((|0\rangle + |1\rangle)(\langle 0| + \langle 1|)\right) + \frac{1}{4}\left((|0\rangle - |1\rangle)(\langle 0| - \langle 1|)\right) \\ &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|, \end{aligned} \tag{1.50}$$

with matrix representation

$$\rho = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{1.51}$$

There is a cute observation that is important at this point: There are many different ways of preparing the same density operator! Since all expectation values of observables are computed as

$$\langle A \rangle = \mathrm{tr}(A\rho), \tag{1.52}$$

we get exactly same same value for all observables in case of

$$\rho = \sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j| = \sum_{k=1}^{m} q_k |\phi_k\rangle\langle\phi_k|, \tag{1.53}$$

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

## 1.2 Measurement postulate

The measurement postulate has to give an answer to the following questions: What are the outcomes of a measurement? What is the probability of obtaining this? What is the state immediately after the measurement? The measurement postulate settles these questions.

**Measurement postulate:** Let $A$ be an observable with spectral decomposition

$$A = \sum_k \lambda_k P_k, \tag{1.54}$$

where

$$P_k = \sum_{\mathrm{EV}\lambda_k} |\psi_k\rangle\langle\psi_k|. \tag{1.55}$$

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

$$p_k = \mathrm{tr}[\rho P_k]. \tag{1.56}$$

The state immediately after the measurement is

$$\rho_k' = \frac{P_k \rho P_k}{\mathrm{tr}[P_k \rho P_k]} = \frac{P_k \rho P_k}{\mathrm{tr}[\rho P_k]}. \tag{1.57}$$

## 1.3 Composite 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

$$\mathcal{H}_1 = \mathrm{span}\{|0\rangle, \ldots, |d_1 - 1\rangle\}. \tag{1.58}$$

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

$$\mathcal{H}_2 = \mathrm{span}\{|0\rangle, \ldots, |d_2 - 1\rangle\}. \tag{1.59}$$

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*

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{1.60}$$

It is spanned by the orthonormal basis vectors

$$\{|j\rangle \otimes |k\rangle : j = 0, \ldots, d_1 - 1; k = 0, \ldots, d_2 - 1\}. \tag{1.61}$$

Such basis elements of tensor products are sometimes also written as

$$\{|j, k\rangle : j = 0, \ldots, d_1 - 1; k = 0, \ldots, d_2 - 1\}. \tag{1.62}$$

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

$$|\psi_1\rangle = \sum_{j=0}^{d_1-1} \alpha_j |j\rangle \tag{1.63}$$

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

$$|\psi_2\rangle = \sum_{j=0}^{d_2-1} \beta_j |j\rangle, \tag{1.64}$$

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

$$|\psi\rangle = \sum_{j=0}^{d_1-1} \sum_{k=0}^{d_2-1} \gamma_{j,k} |j\rangle \otimes |k\rangle, \tag{1.65}$$

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

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{1.66}$$

Similarly, an arbitrary linear operator can be decomposed as

$$O = \sum_{j,k} c_{j,k} A_j \otimes B_k, \tag{1.67}$$

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

$$\mathcal{H} \otimes \cdots \otimes \mathcal{H} = \mathcal{H}^{\otimes n}. \tag{1.68}$$

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

$$\mathcal{H} = L^2(\mathbb{R}) \otimes \mathbb{C}^2. \tag{1.69}$$

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

$$\mathcal{H} = L^2(\mathbb{R}) \otimes L^2(\mathbb{R}) \otimes L^2(\mathbb{R}). \tag{1.70}$$

## 1.4 Time evolution

When we prepare a quantum system in some state $\rho$, how does it evolve in time? The answer to this question is given by the Schrödinger equation. It is given in the form of a differential equation.

**Schrödinger equation:** State vectors of pure states evolve in time according to

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle, \tag{1.71}$$

where $H$ is the *Hamiltonian*. General quantum states evolve as

$$i\hbar\frac{\partial}{\partial t}\rho(t) = [H, \rho(t)]. \tag{1.72}$$

Making use of the unitary *time evolution operator*

$$U(t) = e^{-iHt/\hbar}, \tag{1.73}$$

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

$$U(t)U^\dagger(t) = U^\dagger(t)U(t) = \mathbb{1}. \tag{1.74}$$

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

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

$$|\psi(t_2)\rangle = U(t_2 - t_1)|\psi(t_1)\rangle. \tag{1.75}$$

General states, so density operators, evolve according to

$$\rho(t_2) = U(t_2 - t_1)\rho(t_1)U^\dagger(t_2 - t_1). \tag{1.76}$$

This type of time evolution is referred to as time evolution in the *Schrödinger* picture, in which observables are kept constant and quantum states evolve. It can also make sense to refer to a picture in which states following preparations are kept constant and observables evolve. This picture is referred to as *Heisenberg picture*.

**Evolution of observables in the Heisenberg picture:** In the Heisenberg picture, observable $A$ evolve from time $t_1$ to $t_2 > t_1$ as

$$A(t_2 - t_1) = U^\dagger(t_2 - t_1) A U(t_2 - t_1). \tag{1.77}$$

Of course, the predictions in both pictures are identical,

$$\begin{aligned} \langle A(t) \rangle &= \text{tr}[A\rho(t)] = \text{tr}[A(U(t)\rho U^\dagger(t))] = \text{tr}[(U^\dagger(t) A U(t))\rho] \\ &= \text{tr}[A(t)\rho]. \end{aligned} \tag{1.78}$$

with $t_2 = t$ und $t_1 = 0$ to simplify the notation.

## 1.5 Heisenberg's uncertainty principle

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 pre-factor 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 $|\psi\rangle$,

$$(\Delta A)^2 = \langle (A - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2. \tag{1.79}$$

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

$$C = A - \langle A \rangle, \tag{1.80}$$

$$D = B - \langle B \rangle. \tag{1.81}$$

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

$$\langle CD \rangle = \langle \psi | CD | \psi \rangle = z = x + iy, \tag{1.82}$$

$$\langle DC \rangle = \langle \psi | (CD)^\dagger | \psi \rangle = z^* = x - iy, \tag{1.83}$$

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

$$\langle [C, D] \rangle = \langle \psi | [C, D] | \psi \rangle = z - z^* = 2iy = 2i \,\text{im}(\langle CD \rangle). \tag{1.84}$$

Taking the absolute value, we obtain

$$|\langle [C, D]\rangle|^2 \leq 4|\langle\psi|CD|\psi\rangle|^2. \tag{1.85}$$

Of course, we have for the commutator

$$[C, D] = [A, B], \tag{1.86}$$

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

$$\begin{aligned} |\langle\psi|CD|\psi\rangle|^2 &= |(\langle\psi|C)\cdot(D|\psi\rangle)|^2 \\ &\leq \|C|\psi\rangle\|^2 \cdot \|D|\psi\rangle\|^2 \\ &= \langle\psi|C^2|\psi\rangle \cdot \langle\psi|D^2|\psi\rangle. \end{aligned} \tag{1.87}$$

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

$$\Delta A \cdot \Delta B \geq \frac{1}{2}|\langle\psi|[A, B]|\psi\rangle|. \tag{1.88}$$

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

$$\Delta X \cdot \Delta P \geq \frac{\hbar}{2}, \tag{1.89}$$

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 pre-factor 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 pre-factor. 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.