CHAPTER ONE

THE QUANTUM MECHANICAL STATE SPACE

Section (1.1): Introduction to the Quantum Mechanical State Space

When we talk about physics, we attempt to find a mathematical description of the world. Of course, such a description cannot be justified from mathematical consistency alone, but has to agree with experimental evidence. The mathematical concepts that are introduced are usually motivated from our experience of nature. Concepts such as position and momentum or the state of a system are usually taken for granted in classical physics. However, many of these have to be subjected to a careful re-examination when we try to carry them over to quantum physics. One of the basic notions for the description of a physical system is that of its 'state'. The 'state' of a physical system essentially can then be defined, roughly, as the description of all the known (in fact one should say knowable) properties of that system and it therefore represents your knowledge about this system. The set of all states forms what we usually call the state space. In classical mechanics for example this is the phase space (the variables are then position and momentum), which is a real vector space. For a classical point-particle moving in one dimension, this space is two dimensional, one dimension for position, one dimension for momentum. We know that the quantum mechanical state space differs from that of classical mechanics. One reason for this can be found in the ability of quantum systems to exist in coherent superposition of states with complex amplitudes, other differences relate to the description of multi-particle systems. This suggests, that a good choice for the quantum mechanical state space may be a complex vector space.

Before we begin to investigate the mathematical foundations of quantum mechanics, we would like to present a simple example (including some live experiments) which motivates the choice of complex vector spaces as state spaces a bit more. Together with the hypothesis of the existence of photons it will allow us also to 'derive', or better, to make an educated guess for the projection postulate and the rules for the computation of measurement outcomes.

In the next subsections we will briefly motivate that the quantum mechanical state space should be a complex vector space and also motivate some of the other postulates of quantum mechanics.

(1.1.1): From Polarized Light to Quantum Theory

Let us consider plane waves of light propagating along the z-axis. This light is described by the electric field vector $\vec{E}$ orthogonal on the direction of propagation. The electric field vector determines the state of light. the magnetic field is given by $\vec{B} = \vec{e}_z \times \vec{E}$. Given the electric and magnetic field, Maxwells equations determine the further time evolution of these fields. In the absence of charges, we know that $\vec{E}(\vec{r}, t)$ cannot have a z-component, so that we can write

$$\vec{E}(\vec{r}, t) = E_x(\vec{r}, t)\vec{e}_x + E_y(\vec{r}, t)\vec{e}_y = \begin{pmatrix} E_x(\vec{r}, t) \\ E_y(\vec{r}, t) \end{pmatrix} \tag{1.1}$$

The electric field is real valued quantity and the general solution of the free wave equation is given by

$$E_x(\vec{r}, t) = E_x^0 \cos(kz - wt + \alpha_x)$$

$$E_y(\vec{r}, t) = E_y^0 \cos(kz - wt + \alpha_y)$$

Here $k = 2\pi/\lambda$ is the wave-number, $w = 2\pi v$ the frequency, α_x and α_y are the real phases and E_x^0 and E_y^0 the real valued amplitudes of the field components. The energy density of the field is given by

$$\epsilon(\vec{r}, t) = \frac{1}{8\pi}\left(\vec{E}^2(\vec{r}, t) + \vec{B}^2(\vec{r}, t)\right)$$

$$= \frac{1}{4\pi}\left[(E_x^0)^2 \cos^2(kz - wt + \alpha_x) + \left(E_y^0\right)^2 \cos^2(kz - wt + \alpha_y)\right]$$

For a fixed position $\vec{r}$ we are generally only really interested in the timeaveraged energy density which, when multiplied with the speed of light, determines the rate at which energy flows in z-direction. Averaging over one period of the light we obtain the averaged energy density $\bar{\epsilon}(\vec{r})$ with

$$\bar{\epsilon}(\vec{r}) = \frac{1}{8\pi}\left[(E_x^0)^2 + \left(E_y^0\right)^2\right] \tag{1.2}$$

For practical purposes it is useful to introduce the complex field components

$$E_x(\vec{r},t) = Re\left(E_x e^{i(kz-wt)}\right) \quad E_y(\vec{r},t) = Re\left(E_y e^{i(kz-wt)}\right) \tag{1.3}$$

With $E_x = E_x^0 e^{i\alpha_x}$ and $E_y = E_y^0 e^{i\alpha_y}$. Comparing with Eq. (1.2) we find that the averaged energy density is given by

$$\bar{\epsilon}(\vec{r}) = \frac{1}{8\pi}\left[|E_x|^2 + |E_y|^2\right] \tag{1.4}$$

Usually one works with the complex field

$$E_x(\vec{r},t) = \left(E_x \vec{e}_x + E_y \vec{e}_y\right)e^{i(kz-wt)} = \begin{pmatrix} E_x \\ E_y \end{pmatrix}e^{i(kz-wt)} \tag{1.5}$$

This means that we are now characterizing the state of light by a vector with complex components.

The polarization of light waves are described by E_x and E_y. In the general case of complex E_x and E_y we will have elliptically polarized light. There are a number of important special cases.

(1) $E_y = 0$: linear polarization along the x-axis.
(2) $E_x = 0$: linear polarization along the y-axis.
(3) $E_x = E_y$: linear polarization along 45^0-axis
(4) $E_y = iE_x$: Right circularly polarized light.
(5) $E_y = -iE_x$: Left circularly polarized light.

In the following we would like to consider some simple experiments for which I will compute the outcomes using classical electrodynamics. Then I will go further and use the hypothesis of the existence of photons to derive a number of quantum mechanical rules from these experiments.

Experiment (1.1.1) :

Let us first consider a plane light wave propagating in z-direction that is falling onto an x-polarizer which allows x-polarized light to pass through (but not

3

y polarized light). After passing the polarizer the light is x-polarized and Light of arbitrary polarization is hitting a x-polarizer.

The expression for the energy density Eq. (1.4) we find that the ratio between incoming intensity I_{in} (energy density times speed of light) and outgoing intensity I_{out} is given by

$$\frac{I_{out}}{I_{in}} = \frac{|E_x|^2}{|E_x|^2 + |E_y|^2} \tag{1.6}$$

So far this looks like an experiment in classical electrodynamics or optics.

Now to discuss the quantum interpretation, let us change the way of looking at this problem and thereby turn it into a quantum mechanical experiment. We heard at various points in our physics course that light comes in little quanta known as photons. The first time this assumption had been made was by Planck in 1900 'as an act of desperation' to be able to derive the blackbody radiation spectrum. Indeed, you can also observe in direct experiments that the photon hypothesis makes sense. When you reduce the intensity of light that falls onto a photodetector, you will observe that the detector responds with individual clicks each triggered by the impact of a single photon (if the detector is sensitive enough). The photo-electric effect and various other experiments also confirm the existence of photons. So, in the low-intensity limit we have to consider light as consisting of indivisible units called photons. It is a fundamental property of photons that they cannot be split – there is no such thing as half a photon going through a polarizer for example. In this photon picture we have to conclude that sometimes a photon will be absorbed in the polarizer and sometimes it passes through. If the photon passes the polarizer, we have gained one piece of information, namely that the photon was able to pass the polarizer and that therefore it has to be polarized in x-direction. The probability p for the photon to pass through the polarizer is obviously the ratio between transmitted and incoming intensities, which is given by

$$p = \frac{|E_x|^2}{|E_x|^2 + |E_y|^2} \tag{1.7}$$

If we write the state of the light with normalized intensity

$$\vec{E}_N = \frac{E_x}{\sqrt{|E_x|^2 + |E_y|^2}} \vec{e}_x + \frac{E_y}{|E_x|^2 + |E_y|^2} \vec{e}_y \tag{1.8}$$

Then in fact we find that the probability for the photon to pass the x-polarizer is just the square of the amplitude in front of the basis vector $\vec{e}_x$.

Furthermore we see that the state of the photon after it has passed the x-polarizer is given by

$$\vec{E}_N = \vec{e}_x \tag{1.9}$$

ie the state has has changed from $\begin{pmatrix} E_x \\ E_y \end{pmatrix}$ to $\begin{pmatrix} E_x \\ 0 \end{pmatrix}$. This transformation of the state can be described by a matrix acting on vectors, ie

$$\begin{pmatrix} E_x \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{1.10}$$

The matrix that we have written here has eigenvalues 0 and 1 and is therefore a projection operator, in fact this reminds strongly of the projection postulate in quantum mechanics.

Experiment (1.1.2) :

Now let us make a somewhat more complicated experiment by placing a second polarizer behind the first x-polarizer. The second polarizer allows photons polarized in x' direction to pass through. If we slowly rotate the polarizer from the x direction to the y direction, we observe that the intensity of the light that passes through the polarizer decreases and vanishes when the directions of the two polarizers are orthogonal. We would like to describe this experiment mathematically. How do we compute the intensity after the polarizer now? To this end we need to see how we can express vectors in the basis chosen by the direction x' in terms of the old basis vectors $\vec{e}_x, \vec{e}_y$.

The new rotated basis $\vec{e}_x, \vec{e}_y$ can be expressed by the old basis by

$$\vec{e}_x = \cos \phi \, \vec{e}_x + \sin \phi \, \vec{e}_y \quad \vec{e}_y = -\sin \phi \, \vec{e}_x + \cos \phi \, \vec{e}_y \tag{1.11}$$

And vice versa

$$\vec{e}_x' = \cos\phi\,\vec{e}_x - \sin\phi\,\vec{e}_y \quad \vec{e}_y' = \sin\phi\,\vec{e}_x + \cos\phi\,\vec{e}_y \tag{1.12}$$

Note that $\cos\phi = \vec{e}_x'.\vec{e}_x$ and $\sin\phi = \vec{e}_y'.\vec{e}_y$ where we have used the real scalar product between vectors.

The state of the x-polarized light after the first polarizer can be rewritten in the new basis of the x'-polarizer. We find

$$\vec{E} = E_x\vec{e}_x = E_x\cos\phi\,\vec{e}_x' - E_x\sin\phi\,\vec{e}_y' = E_x(\vec{e}_x'.\vec{e}_x)\vec{e}_x' - E_x(\vec{e}_y'.\vec{e}_y)\vec{e}_y'$$

Now we can easily compute the ratio between the intensity before and after the x'-polarizer. We find that it is

$$\frac{I_{after}}{I_{before}} = |\vec{e}_x'.\vec{e}_x|^2 = \cos^2\phi \tag{1.13}$$

Or if we describe the light in terms of states with normalized intensity as in equation (1.8), then we find that

$$\frac{I_{after}}{I_{before}} = |\vec{e}_x'.\vec{E}_N|^2 = \frac{|\vec{e}_x'.\vec{E}_N|^2}{|\vec{e}_x'.\vec{E}_N|^2 + |\vec{e}_y'.\vec{E}_N|^2} \tag{1.14}$$

Where $\vec{E}_N$ is the normalized intensity state of the light after the x-polarizer.

This demonstrates that the scalar product between vectors plays an important role in the calculation of the intensities (and therefore the probabilities in the photon picture).

Varying the angle ϕ between the two bases we can see that the ratio of the incoming and outgoing intensities decreases with increasing angle between the two axes until the angle reaches 90^0 degrees.

Interpretation(1.1.3) :

Viewed in the photon picture this is a rather surprising result, as we would have thought that after passing the x-polarizer the photon is 'objectively' in the x-polarized state. However, upon probing it with an x'-polarizer we find that it also has a quality of an x'-polarized state. In the next experiment we will see an even more worrying result. For the moment we note that the state of a photon

can be written in different ways and this freedom corresponds to the fact that in quantum mechanics we can write the quantum state in many different ways as a quantum superpositions of basis vectors.

Let us push this idea a bit further by using three polarizers in a row.

Experiment (1.1.4) :

If after passing the x-polarizer, the light falls onto a y-polarizer, then no light will go through the polarizer because the two directions are perpendicular to each other. This simple experimental result changes when we place an additional polarizer between the x and the y-polarizer. Assume that we place a x'-polarizer between the two polarizers. Then we will observe light after the y-polarizer depending on the orientation of x'. The light after the last polarizer is described by $\tilde{E}_y \vec{e}_y$. The amplitude ˜Ey is calculated analogously as in Experiment (1.1.2). Now let us describe the (x-x'-y) experiment mathematically. The complex electric field (without the time dependence) is given by

before the x-polarizer:

$$\vec{E}_1 = E_x \vec{e}_x + E_y \vec{e}_y$$

after the x-polarizer:

$$\vec{E}_2 = \left(\vec{E}_1 \vec{e}_x\right)\vec{e}_x = E_x \vec{e}_x = E_x \cos\phi\, \vec{e}_x - E_x \sin\phi\, \vec{e}_y$$

after the x'-polarizer:

$$\vec{E}_3 = \left(\vec{E}_2 \vec{e}_x\right)\vec{e}_x = E_x \cos\phi\, \vec{e}_x = E_x \cos^2\phi\, \vec{e}_x + E_x \cos\phi \sin\phi\, \vec{e}_y$$

after the y-polarizer:

$$\vec{E}_4 = \left(\vec{E}_3 \vec{e}_y\right)\vec{e}_y = E_x \cos\phi \sin\phi\, \vec{e}_y$$

Therefore the ratio between the intensity before the x'-polarizer and after the y-polarizer is given by

$$\frac{I_{after}}{I_{before}} = \cos^2\phi \sin^2\phi \tag{1.15}$$

7

Again, if we interpret this result in the photon picture, then we arrive at the conclusion, that the probability for the photon to pass through both the x' and the y polarizer is given by $\cos^2\phi\sin^2\phi$. This experiment further highlights the fact that light of one polarization may be interpreted as a superposition of light of other polarizations. This superposition is represented by adding vectors with complex coefficients. If we consider this situation in the photon picture we have to accept that a photon of a particular polarization can also be interpreted as a superposition of different polarization states.

All these observations suggest that complex vectors, their amplitudes, scalar products and linear transformations between complex vectors are the basic ingredient in the mathematical structure of quantum mechanics as opposed to the real vector space of classical mechanics. Therefore the rest of this chapter will be devoted to a more detailed introduction to the structure of complex vector-spaces and their properties.

(1.1.2): Complex vector spaces

We will now give a formal definition of a complex vector space and will then present some of its properties. Before we come to this definition, we introduce a standard notation that we will use in this chapter. Given some set V we define

Notation (1.1.5) :

(1) $\forall |x\rangle \in V$ means: For all $|x\rangle$ that lie in V.
(2) $\exists |x\rangle \in V$ *means: There exists an element $|x\rangle$ that lies in V.*

Note that we have used a somewhat unusual notation for vectors. We have replaced the vector arrow on top of the letter by a sort of bracket around the letter. We will use this notation when we talk about complex vectors, in particular when we talk about state vectors. Now we can state the definition of the complex vector space.

Definition (1.1.6) :

Given a $(V, C, +, \cdot)$ where V is a set of objects (usually called vectors), C denotes the set of complex numbers, + denotes the group operation of addition

and $\cdot$ denotes the multiplication of a vector with a complex number. $(V, C, +, \cdot)$ is called a complex vector space if the following properties are satisfied:

(1) $(V, +)$ *is an* Abelian group, which means that

    a. $\forall |a\rangle, |b\rangle \in V \Rightarrow |a\rangle + |b\rangle \in V.$ (closure)

    b. $\forall |a\rangle, |b\rangle, |c\rangle \in V \Rightarrow |a\rangle + |b\rangle + |c\rangle = (|a\rangle + |b\rangle) + |c\rangle.$ (associative)

    c. $\exists |O\rangle \in V$ *so that* $\forall |a\rangle \in V \Rightarrow |a\rangle + |O\rangle = |a\rangle.$ (zero)

    d. $\forall |a\rangle, |b\rangle \in V: \exists(-|a\rangle) \in V$ *so that* $|a\rangle + (-|a\rangle) = |O\rangle.$ (inverse)

    e. $\forall |a\rangle, |b\rangle \in V \Rightarrow |a\rangle + |b\rangle = |b\rangle + |a\rangle.$ (Abelian)

(2) The Scalar multiplication satisfies

    a. $\forall \alpha \in C, |x\rangle \in V$ $\Rightarrow \alpha|x\rangle \in V$

    b. $\forall |x\rangle \in V$ $\Rightarrow 1.|x\rangle = |x\rangle$ (unit)

    c. $\forall c, d \in C, |x\rangle \in V$ $\Rightarrow (c.d).|x\rangle = c.(d.|x\rangle)$ (associative)

    d. $\forall c, d \in C, |x\rangle, |y\rangle \in V$ $\Rightarrow c.(|x\rangle + c.|y\rangle) = c.|x\rangle + c.|y\rangle$

    and $(c + d).|x\rangle = c.|x\rangle + c.|y\rangle$ (distributive)

This definition looks quite abstract but a few examples will make it clearer.

Examples (1.1.7) :

  (1) **A simple proof**

We would like to show how to prove the statement $0.|x\rangle = |O\rangle$. This might look trivial, but nevertheless we need to prove it, as it has not been stated as an axiom. From the axioms given in Definition (1.1.6). We conclude.

$$|O\rangle \overset{(1d)}{=} -|x\rangle + |x\rangle$$

$$\overset{(2d)}{=} -|x\rangle + 1.|x\rangle$$

$$= -|x\rangle + (1 + 0).|x\rangle$$

$$\overset{(2d)}{=} -|x\rangle + 1.|x\rangle + 0.|x\rangle$$

$$\overset{(2d)}{=} -|x\rangle + |x\rangle + 0.|x\rangle$$

$$\overset{(1d)}{=} |0\rangle + 0.|x\rangle$$

$$\overset{(1c)}{=} 0.|x\rangle$$

(2) The C^2

This is the set of two-component vectors of the form

$$|a\rangle = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{1.16}$$

Where the a_i are complex numbers. The addition and scalar multiplication are defined as

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} := \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix} \tag{1.17}$$

$$c.\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} := \begin{pmatrix} c.a_1 \\ c.a_2 \end{pmatrix} \tag{1.18}$$

It is now easy to check that $V = C^2$ together with the addition and scalar multiplication defined above satisfy the definition of a complex vector space. The vector space C^2 is the one that is used for the description of spin-$\frac{1}{2}$ particles such as electrons.

(3) The set of real functions of one variable $f : R \rightarrow R$

The group operations are defined as

$$(f_1 + f_2)(x) := f_1(x) + f_2(x)$$

$$(c.f)(x) = c.f(x)$$

Again it is easy to check that all the properties of a complex vector space are satisfied.

(4) Complex $n \times n$ matrices

The elements of the vector space are

$$M = \begin{pmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{nn} \end{pmatrix} \tag{1.19}$$

Where the m_{ij} are arbitrary complex numbers. The addition and scalar multiplication are defined as

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{nn} + b_{nn} \end{pmatrix}$$

$$c. \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} c.\,a_{11} & \cdots & c.\,a_{1n} \\ \vdots & \ddots & \vdots \\ c.\,a_{n1} & \cdots & c.\,a_{nn} \end{pmatrix}$$

Again it is easy to confirm that the set of complex $n \times n$ matrices with the rules that we have defined here forms a vector space. Note that we are used to consider matrices as objects acting on vectors, but as we can see here we can also consider them as elements (vectors) of a vector space themselves.

(1.1.3): Basis and Dimension

Some of the most basic concepts of vector spaces are those of linear independence, dimension and basis. They will help us to express vectors in terms of other vectors and are useful when we want to define operators on vector spaces which will describe observable quantities.

Quite obviously some vectors can be expressed by linear combinations of others. For example

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2.\begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.20}$$

It is natural to consider a given set of vectors $\{|x\rangle_1, \ldots, |x\rangle_k\}$ and to ask the question, whether a vector in this set can be expressed as a linear combination of the others. Instead of answering this question directly we will first consider a slightly different question. Given a set of vectors $\{|x\rangle_1, \ldots, |x\rangle_k\}$, can the null vector $|0\rangle$ can be expressed as a linear combination of these vectors? This means that we are looking for a linear combination of vectors of the form

$$\lambda_1 |x\rangle_1 + \ldots + \lambda_2 |x\rangle_k = |0\rangle \tag{1.21}$$

Clearly Eq. (1.21) can be satisfied when all the λ_i vanish. But this case is trivial and we would like to exclude it. Now there are two possible cases left:

(a) There is no combination of λ_i's, not all of which are zero, that satisfies Eq.(1.21).

(b) There are combinations of λ_i's, not all of which are zero, that satisfy Eq. (1.21).

These two situations will get different names and are worth,

Definition (1.1.8) :

A set of vectors $\{|x\rangle_1, \ldots, |x\rangle_k\}$ is called linearly independent if the equation

$$\lambda_1|x\rangle_1 + \ldots + \lambda_2|x\rangle_k = |\mathcal{O}\rangle \tag{1.22}$$

has only the trivial solution $\lambda_1 = \ldots = \lambda_k = 0$.

If there is a nontrivial solution to Eq. (1.22), i.e. at least one of the $\lambda_i = 0$, then we call the vectors $\{|x\rangle_1, \ldots, |x\rangle_k\}$ linearly dependent.

Now we are coming back to our original question as to whether there are vectors in $\{|x\rangle_1, \ldots, |x\rangle_k\}$ that can be expressed by all the other vectors in that set. As a result of this definition we can see the following

Lemma (1.1.9) :

For a set of linearly independent vectors $\{|x\rangle_1, \ldots, |x\rangle_k\}$, no $|x\rangle_i$ can be expressed as a linear combination of the other vectors, i.e. one cannot find λ_j that satisfy the equation

$$\{\lambda_1|x\rangle_1 + \cdots + \lambda_{i-1}|x\rangle_{i-1} + \cdots + \lambda_{i+1}|x\rangle_{i+1} + \cdots + \lambda_k|x\rangle_k = |x\rangle_i\} \tag{1.23}$$

In a set of linearly dependent vectors $\{|x\rangle_1, \ldots, |x\rangle_k\}$ there is at least one $|x\rangle_i$ that can be expressed as a linear combination of all the other $|x\rangle_j$.

Example (1.1.10) :

The set $\{|\mathcal{O}\rangle\}$ consisting of the null vector only, is linearly dependent in a sense that will become clearer when we really talk about quantum mechanics, in

a set of linearly independent set of vectors, each vector has some quality that none of the other vectors have.

After we have introduced the notion of linear dependence, we can now proceed to define the dimension of a vector space. Evidently a plain surface is 2-dimensional and space is 3-dimensional. Why do we say this? Consider a plane, for example. Clearly, every vector in the plane can be expressed as a linear combination of any two linearly independent vectors $|e\rangle_1, |e\rangle_2$. As a result you will not be able to find a set of three linearly independent vectors in a plane, while two linearly independent vectors can be found. This is the reason to call a plane a two-dimensional space. Let's formalize this observation in the following definition

Definition (1.1.11) :

The dimension of a vector space V is the largest number of linearly independent vectors in V that one can find.

Now we introduce the notion of basis of vector spaces.

Definition (1.1.12) :

A set of vectors $\{|x\rangle_1, \ldots, |x\rangle_k\}$ is called a basis of a vector space V if

(a) $\{|x\rangle_1, \ldots, |x\rangle_k\}$ are linearly independent.
(b) $\forall |x\rangle \in V : \exists \lambda_i \in C \Longrightarrow x = \sum_{i=1}^{k} \lambda_i |x\rangle_i$

Condition (b) states that it is possible to write every vector as a linear combination of the basis vectors. The first condition makes sure that the set $\{|x\rangle_1, \ldots, |x\rangle_k\}$ is the smallest possible set to allow for condition (b) to be satisfied. It turns out that any basis of an N dimensional vector space V contains exactly N vectors. Let us illustrate the notion of basis.

Examples (1.1.13) :

(1) Consider the space of vectors C^2 with two components. Then the two vectors

$$|x\rangle_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |x\rangle_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.24}$$

Form a basis of C^2. A basis for the C^N can easily be constructed in the same way.

(2) An example for an infinite dimensional vector space is the space of complex polynomials, i.e. the set

$$V = \{c_0 + c_1 z + \cdots + c_k z^k \,|\, \text{arbitrary k and } \forall c_i \in C \qquad (1.25)$$

Two polynomials are equal when they give the same values for all $z \in C$. Addition and scalar multiplication are defined coefficient wise. It is easy to see that the set $\{1, z, z^2, \dots\}$ is linearly independent and that it contains infinitely many elements. Together with other examples you will prove (in the problem sheets) that Eq. (1.25) indeed describes a vector space.

(1.1.4): Scalar Products and Norms on Vector Spaces

Any set of N linearly independent vectors of an N dimensional vector space V form a basis. But not all such choices are equally convenient. To find useful ways to chose a basis and to find a systematic method to find the linear combinations of basis vectors that give any arbitrary vector $|x\rangle \in V$ we will now introduce the concept of scalar product between two vectors. This is not to be confused with scalar multiplication which deals with a complex number and a vector. The concept of scalar product then allows us to formulate what we mean by orthogonality. Subsequently we will define the norm of a vector, which is the abstract formulation of what we normally call a length. This will then allow us to introduce orthonormal bases which are particularly handy.

The scalar product will play an extremely important role in quantum mechanics as it will in a sense quantify how similar two vectors (quantum states) are. you can easily see qualitatively that the pairs of vectors become more and more different from left to right. The scalar product puts this into a quantitative form. This is the reason why it can then be used in quantum mechanics to quantify how likely it is for two quantum states to exhibit the same behaviour in an experiment.

To introduce the scalar product we begin with an abstract formulation of the properties that we would like any scalar product to have. Then we will have a look at examples which are of interest for quantum mechanics.

Definition (1.1.14) :

A complex scalar product on a vector space assigns to any two vectors $|x\rangle, |y\rangle \in V$ a complex number $(|x\rangle, |y\rangle) \in C$ satisfying the following rules

(a) $\forall |x\rangle, |y\rangle, |z\rangle \in V, \alpha_i \in C: (|x\rangle, \alpha_1|y\rangle, \alpha_2|z\rangle) = \alpha_1(|x\rangle, |y\rangle) + \alpha_2(|x\rangle, |z\rangle)$.

(b) $\forall |x\rangle, |y\rangle \in V : (|x\rangle, |y\rangle) = (|x\rangle, |y\rangle)^*$. *(symmetry)*

(c) $\forall |x\rangle \in V: (|x\rangle, |y\rangle) \geq 0$. *(positivity)*

(d) $\forall |x\rangle \in V: (|x\rangle, |y\rangle) = 0 \Longleftrightarrow |x\rangle = |O\rangle$.

These properties are very much like the ones that you know from the ordinary dot product for real vectors, except for property 2 which we had to introduce in order to deal with the fact that we are now using complex numbers. Note that we only defined linearity in the second argument. This is in fact all we need to do.

$$\forall |x\rangle, |y\rangle, |z\rangle \in V, \alpha \in C: (\alpha|x\rangle, \beta|y\rangle, |z\rangle) = \alpha^*(|x\rangle, |z\rangle) + \beta^*(|y\rangle, |z\rangle) \quad (1.26)$$

Vector spaces on which we have defined a scalar product are also called unitary vector spaces. We will now present some examples that play significant roles in quantum mechanics.

Examples (1.1.15) :

(1) **The scalar product in C^n**

Given two complex vectors $|x\rangle, |y\rangle \in C^n$ with components x_i and y_i we define the scalar product

$$|x\rangle, |y\rangle = \sum_{i=1}^{n} x_i^* y_i \quad (1.27)$$

Where $*$ denotes the complex conjugation.

(2) **Scalar product on continuous square integrable functions**

A square integrable function $\psi \in \mathcal{L}^2(R)$ is one that satisfies

$$\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx < \infty \quad (1.28)$$

Eq. (1.28) already implies how to define the scalar product for these square integrable functions. For any two functions , $\psi, \phi \in \mathcal{L}^2(\boldsymbol{R})$ we define

$$(\psi, \phi) = \int_{-\infty}^{\infty} \psi(x)^* \phi(x) \, dx \qquad (1.29)$$

We can even define the scalar product for discontinuous square integrable functions, but then we need to be careful when we are trying to prove property (4) for scalar products. One reason is that there are functions which are nonzero only in isolated points (such functions are discontinuous) and for which Eq. (1.28) vanishes. An example is the function

$$f(x) = \begin{cases} 1 \text{ for } x = 0 \\ 0 \text{ anywhere else} \end{cases}$$

The solution to this problem lies in a redefinition of the elements of our set. If we identify all functions that differ from each other only in countably many points then we have to say that they are in fact the same element of the set. If we use this redefinition then we can see that also condition 4 of a scalar product is satisfied.

An extremely important property of the scalar product is the Schwarz inequality which is used in many proofs. In particular I will used it to prove the triangular inequality for the length of a vector and in the proof of the uncertainty principle for arbitrary observables.

Theorem (1.1.16): (The Schwarz inequality)

For any $|x\rangle, |y\rangle \in V$ we have

$$|(|x\rangle, |y\rangle)|^2 \leq (|x\rangle, |x\rangle)(|y\rangle, |y\rangle) \qquad (1.30)$$

Proof

For any complex number α we have

$$0 \leq (|x\rangle + \alpha|y\rangle, |x\rangle + \alpha|y\rangle)$$

$$= (|x\rangle, |x\rangle) + \alpha(|x\rangle, |y\rangle) + \alpha^*(|y\rangle, |x\rangle) + |\alpha|^2(|y\rangle, |y\rangle)$$

$$= (|x\rangle, |x\rangle) + 2vRe(|x\rangle, |y\rangle) - 2wIm(|x\rangle, |y\rangle) + (v^2 + w^2)(|y\rangle, |y\rangle)$$

$$=: f(v, w) \tag{1.31}$$

In the definition of $f(v, w)$ in the last row we have assumed $\alpha = v + iw$. To obtain the sharpest possible bound in Eq. (1.30), we need to minimize the right hand side of Eq. (1.31). To this end we calculate

$$0 = \frac{df}{dv}(u, w) = 2Re(|x\rangle, |y\rangle) + 2v(|y\rangle, |y\rangle) \tag{1.32}$$

$$0 = \frac{df}{dw}(u, w) = -2Im(|x\rangle, |y\rangle) + 2w(|y\rangle, |y\rangle) \tag{1.33}$$

Solving these equations, we find

$$\alpha_{min} = v_{min} + iw_{min} = -\frac{Re(|x\rangle, |y\rangle) - iIm(|x\rangle, |y\rangle)}{(|y\rangle, |y\rangle)} = -\frac{(|y\rangle, |x\rangle)}{(|y\rangle, |y\rangle)} \tag{1.34}$$

Because all the matrix of second derivatives is positive definite, we really have a minimum. If we insert this value into Eq. (1.31) we obtain

$$0 \leq (|x\rangle, |x\rangle) - \frac{(|y\rangle, |x\rangle)(|x\rangle, |y\rangle)}{(|y\rangle, |y\rangle)} \tag{1,35}$$

This implies then Eq. (1.30). Note that we have equality exactly if the two vectors are linearly dependent, i.e. if $|x\rangle = \gamma |y\rangle$.

Having defined the scalar product, we are now in a position to define what we mean by orthogonal vectors.

Definition (1.1.17):

Two vectors $|x\rangle, |y\rangle \in V$ are called orthogonal if

$$(|x\rangle, |y\rangle) = 0 \tag{1.36}$$

We denote with $|x\rangle_{\perp}$ a vector that is orthogonal to $|x\rangle$.

Now we can define the concept of an orthogonal basis which will be very useful in finding the linear combination of vectors that give $|x\rangle$.

Definition (1.1.18):

An orthogonal basis of an N dimensional vector space V is a set of N linearly independent vectors such that each pair of vectors are orthogonal to each other

Example (1.1.19):

In C the three vectors

$$\begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \tag{1.37}$$

Form an orthogonal basis.

Now let us chose an orthogonal basis $\{|x\rangle_1, \ldots, |x\rangle_N\}$ of an N dimensional vector space. For any arbitrary vector $|x\rangle \in V$ we would like to find the coefficients $\lambda_1, \ldots, \lambda_N$ such that

$$\sum_{i=1}^{N} \lambda_i |x\rangle_i \tag{1.38}$$

Of course we can obtain the λ_i by trial and error, but we would like to find an efficient way to determine the coefficients λ_i. To see this, let us consider the scalar product between $|x\rangle$ and one of the basis vectors $|x\rangle_i$. Because of the orthogonality of the basis vectors, we find

$$(|x\rangle_i, |x\rangle) = \lambda_i(|x\rangle_i, |x\rangle_i) \tag{1.39}$$

Note that this result holds true only because we have used an orthogonal basis. Using Eq. (1.39) in Eq. (1.38), we find that for an orthogonal basis any vector $|x\rangle$ can be represented as

$$|x\rangle = \sum_{i=1}^{N} \frac{(|x\rangle_i, |x\rangle)}{(|x\rangle_i, |x\rangle_i)} |x\rangle_i \tag{1.40}$$

In Eq. (1.40) we have the denominator $(|x\rangle_i, |x\rangle)$ which makes the formula a little bit clumsy. This quantity is the square of what we usually call the length of a

vector which shows a vector $|x\rangle = \begin{pmatrix} a \\ b \end{pmatrix}$ in the two-dimensional real vector space $\mathbf{R}^2$. Clearly its length is $\sqrt{a^2 + b^2}$. If we multiply the vector by a number α then we have the vector $\alpha|x\rangle$ which evidently has the length $\sqrt{\alpha^2 a^2 + \alpha^2 b^2} = |\alpha|\sqrt{a^2 + b^2}$. Finally we know that we have a triangular inequality. This means that given two vectors $|x\rangle_1 = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$ and $|x\rangle_2 = \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}$ the length of the $|x\rangle_1 + |x\rangle_2$ is smaller than the sum of the lengths of $|x\rangle_1$ and $|x\rangle_2$. In the following we formalize the concept of a length and we will arrive at the definition of the norm of a vector $|x\rangle_i$. The concept of a norm is important if we want to define what we mean by two vectors being close to one another. In particular, norms are necessary for the definition of convergence in vector spaces, a concept that we will introduce in the next subsection. In the following we specify what properties a norm of a vector should satisfy.

Definition (1.1.20) :

A norm on a vector space V associates with every $|x\rangle \in$ V a real number $\||x\rangle\|$, with the properties.

(1) $\forall |x\rangle \in V: \quad \||x\rangle\| \geq 0 \; and \; \||x\rangle\| = 0 \iff |x\rangle = |\mathcal{O}\rangle.$ *(positivity)*

(2) $\forall |x\rangle \in V, \alpha \in C: \|\alpha|x\rangle\| = |\alpha|.\||x\rangle\|$ *(linearity)*

(3) $\forall |x\rangle, |y\rangle \in V: \||x\rangle + |y\rangle\| \leq \||x\rangle\| + \||y\rangle\|.$ *(triangular inequality)*

A vector space with a norm defined on it is also called a normed vector space. The three properties in Definition (1.1.20) are those that you would intuitively expect to be satisfied for any decent measure of length. As expected norms and scalar products are closely related. In fact, there is a way of generating norms very easily when you already have a scalar product.

Lemma (1.1.21) :

Given a scalar product on a complex vector space, we can define the norm of a vector $|x\rangle$ by

$$\||x\rangle\| = \sqrt{(|x\rangle, |x\rangle)} \qquad\qquad (1.41)$$

Proof

- Properties (1) and (2) of the norm follow almost trivially from the four basic conditions of the scalar product.
- The proof of the triangular inequality uses the Schwarz inequality.

$$
\begin{aligned}
\||x\rangle + |y\rangle\|^2 &= |(|x\rangle + |y\rangle, |x\rangle + |y\rangle)| \\
&= |(|x\rangle, |x\rangle + |y\rangle) + (|y\rangle, |x\rangle + |y\rangle)| \\
&\leq |(|x\rangle, |x\rangle + |y\rangle)| + |(|y\rangle, |x\rangle + |y\rangle)| \\
&\leq \||x\rangle\| \cdot \| |x\rangle + |y\rangle\| + \||y\rangle\| \cdot \||x\rangle + |y\rangle\|
\end{aligned}
\tag{1.42}
$$

Dividing both sides by $\||x\rangle + |y\rangle\|$ yields the inequality. This assumes that the sum $|x\rangle + |y\rangle \neq |0\rangle$. If we have $|x\rangle + |y\rangle = |0\rangle$ then the Schwarz inequality is trivially satisfied.

Lemma (1.1.21) shows that any unitary vector space can canonically (this means that there is basically one natural choice) turned into a normed vector space. The converse is, however, not true. Not every norm gives automatically rise to a scalar product.

Using the concept of the norm we can now define an orthonormal basis for which Eq. (1.40) can then be simplified.

Definition (1.1.22) :

An orthonormal basis of an N dimensional vector space is a set of N pairwise orthogonal linearly independent vectors $\{|x\rangle_1, \ldots, |x\rangle_N\}$ where each vector satisfies $\||x\rangle\|^2 = (|x\rangle_i, |x\rangle_i) = 1$, i.e. they are unit vectors. For an orthonormal basis and any vector $|x\rangle$ we have

$$
|x\rangle = \sum_{i=1}^{N} (|x\rangle_i, |x\rangle) |x\rangle_i = \sum_{i=1}^{N} \alpha_i |x\rangle_i
\tag{1.43}
$$

Where the components of $|x\rangle$ with respect to the basis $\{|x\rangle_1, \ldots, |x\rangle_N\}$ are the $\alpha_i = (|x\rangle_i, |x\rangle)$.

(1.1.5): Completeness and Hilbert Spaces

What do we mean by complete? To see this, let us consider sequences of elements of a vector space (or in fact any set, but we are only interested in vector spaces). We will write sequences in two different ways

$$\{|x\rangle_i\}_i = 0, \dots, \infty \equiv (|x\rangle_0 , |x\rangle_1 , |x\rangle_2 \dots) \tag{1.44}$$

To define what we mean by a convergent sequences, we use norms because we need to be able to specify when two vectors are close to each other.

Definition (1.1.23) :

A sequence $\{|x\rangle_i\}_i = 0, \dots, \infty$ of elements from a normed vector space V converges towards a vector $|x\rangle \in$ V if for all $\epsilon > 0$ there is an n_0 such that for all $n > n_0$ we have

$$\||x\rangle - |x\rangle_n\| \le \epsilon \tag{1.45}$$

But sometimes you do not know the limiting element, so you would like to find some other criterion for convergence without referring to the limiting element. This idea led to the following

Definition (1.1.24) :

A sequence $\{|x\rangle_i\}_i = 0, \dots, \infty$ of elements from a normed vector space V is called a Cauchy sequence if for all $\epsilon > 0$ there is an n_0 such that for all $m, n > n_0$ we have

$$\||x\rangle_m - |x\rangle_n\| \le \epsilon \tag{1.46}$$

Now we can wonder whether every Cauchy sequence converges. Well, it sort of does. But unfortunately sometimes the limiting element does not lie in the set from which you draw the elements of your sequence. How can that be? To illustrate this we will present a vector space that is not complete. Consider the set

$V = \{|x\rangle : \text{only finitely many components of } |x\rangle \text{ are non-zero }\}.$

An example for an element of V is $|x\rangle = (1, 2, 3, 4, 5, 0, \dots)$. It is now quite easy to check that V is a vector-space when you define addition of two vectors via

$$|x\rangle + |y\rangle = (x_1 + y_1 , x_2 + y_2)$$

and the multiplication by a scalar via

$$c|x\rangle = (cx_1 , cx_2 , \dots)$$

Now we define a scalar product from which we will then obtain a norm via the construction of Lemma (1.1.21). We define the scalar product as

$$(|x\rangle, |y\rangle) = \sum_{k=1}^{\infty} x_k^* y_k$$

Now let us consider the series of vectors

$$|x\rangle_1 = (1,0,0,0,\dots)$$

$$|x\rangle_2 = \left(1, \frac{1}{2}, 0, \dots\right)$$

$$|x\rangle_3 = \left(1, \frac{1}{2}, \frac{1}{4}, 0, \dots\right)$$

$$|x\rangle_4 = \left(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, 0, \dots\right)$$

$$\vdots$$

$$|x\rangle_k = (1, \frac{1}{2}, \dots, \frac{1}{2^{k-1}}, 0, \dots)$$

For any n_0 we find that for $m > n > n_0$ we have

$$\||x\rangle_m - |x\rangle_n\| = \left\||0, \dots, 0, \frac{1}{2^n}, \dots, \frac{1}{2^{m-1}}, 0, \dots\right\| \leq \frac{1}{2^{n-1}}$$

Therefore it is clear that the sequence $\{|x\rangle_k\}_{k=1,\dots,\infty}$ is a Cauchy sequence. However, the limiting vector is not a vector from the vector space , because the limiting vector contains infinitely many nonzero elements.

Considering this example let us define what we mean by a complete vector space.

Definition (1.1.25) :

A vector space V is called complete if every Cauchy sequence of elements from the vector space V converges towards an element of V

Now we come to the definition of Hilbert spaces.

Definition (1.1.26) :

A vector space $\mathcal{H}$ is a Hilbert space if it satisfies the following two conditions

(1) $\mathcal{H}$ is a unitary vector space.
(2) $\mathcal{H}$ is complete.

Following our discussions of the vectors spaces, we are now in the position to formulate the first **postulate** of quantum mechanics.

Postulate 1 *The state of a quantum system is described by a vector in a Hilbert space $\mathcal{H}$*

Let us argue physically. We know that we need to be able to represent superpositions, i.e. we need to have a vector space. From the superposition principle we can see that there will be states that are not orthogonal to each other. That means that to some extent one quantum state can be 'present' in another non-orthogonal quantum state -they 'overlap'. The extent to which the states overlap can be quantified by the scalar product between two vectors. In the first section we have also seen, that the scalar product is useful to compute probabilities of measurement outcomes. This requires that we have a norm which can be derived from a scalar product. Because of the obvious usefulness of the scalar product, we require that the state space of quantum mechanics is a vector space equipped with a scalar product. The reason why we demand completeness, can be seen from a physical argument which could run as follows: Consider any sequence of physical states that is a Cauchy sequence. Quite obviously we would expect this sequence to converge to a physical state. It would be extremely strange if by means of such a sequence we could arrive at an unphysical state. Imagine for example that we change a state by smaller and smaller amounts and then suddenly we would arrive at an unphysical state. That makes no sense! Therefore it seems reasonable to demand that the physical state space is complete.

What we have basically done is to distill the essential features of quantum mechanics and to find a mathematical object that represents these essential features without any reference to a special physical system.

In the next subsection we will continue this programme to formulate more principles of quantum mechanics.

(1.1.6): Dirac Notation :

In the following we will introduce a useful way of writing vectors. This notation, the Dirac notation, applies to any vector space and is very useful, in particular it makes life a lot easier in calculations. As most quantum mechanics books are written in this notation it is quite important that you really learn how to use this way of writing vectors. If it appears a bit weird to you in the first place you should just practice its use until you feel confident with it. for example, is to rewrite in Dirac notation all the results that we have presented so far.

So far we have always written a vector in the form $|x\rangle$. The scalar product between two vectors has then been written as $(|x\rangle, |y\rangle)$. Let us now make the following identification

$$|x\rangle \leftrightarrow |x\rangle \tag{1.47}$$

We call $|x\rangle$ a ket. So far this is all fine and well. It is just a new notation for a vector. Now we would like to see how to rewrite the scalar product of two vectors. To understand this best, we need to talk a bit about linear functions of vectors.

Definition (1.1.27) :

A function $f : V \rightarrow C$ from a vector space into the complex numbers is called linear if for any $|\psi\rangle, |\phi\rangle \in V$ and any $\alpha, \beta \in C$ we have

$$f(\alpha|\psi\rangle + \beta|\phi\rangle) = \alpha f(|\psi\rangle) + \beta f(|\phi\rangle) \tag{1.48}$$

With two linear functions f_1, f_2 also the linear combination $\mu f_1 + \upsilon f_2$ is a linear function. Therefore the linear functions themselves form a vector space and it is even possible to define a scalar product between linear functions. The space of the linear function on a vector space V is called the dual space V^*.

Now we would like to show you an example of a linear function which we define by using the scalar product between vectors. We define the function $f_{|\phi\rangle} : V \rightarrow C$, where $|\phi\rangle \in V$ is a fixed vector so that for all $|\psi\rangle \in V$

$$f_{|\phi\rangle}(|\psi\rangle) := (|\phi\rangle, |\psi\rangle) \tag{1.49}$$

Now we would like to introduce a new notation for $f_{|\phi\rangle}$. From now on we will identify

$$f_{|\phi\rangle} \leftrightarrow |\phi\rangle \tag{1.50}$$

and use this to rewrite the scalar product between two vectors $|\phi\rangle, |\psi\rangle$ as

$$\langle\phi|\psi\rangle := \langle\phi(\psi)\rangle \equiv (|\phi\rangle, |\psi\rangle) \tag{1.51}$$

The object $|\phi\rangle$ is called bra and the Dirac notation is therefore sometimes called braket notation.

(1.2): Linear Operators

So far we have only dealt with the elements (vectors) of vector spaces. Now we need to learn how to transform these vectors, that means how to transform one set of vectors into a different set. Again as quantum mechanics is a linear theory we will concentrate on the description of linear operators.

Definition (1.2.1): (Dirac notation)

A linear operator $\widehat{A} : \mathcal{H} \rightarrow \mathcal{H}$ associates to every vector $|\psi\rangle \in \mathcal{H}$ a vector $\widehat{A}|\psi\rangle \in \mathcal{H}$ such that

$$\widehat{A}(\lambda|\psi\rangle + \mu|\phi\rangle) = \lambda\widehat{A}|\psi\rangle + \mu\widehat{A}|\phi\rangle \tag{1.52}$$

for all $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ and $\lambda, \mu \in \mathbb{C}$.

A linear operator $\hat{A} : \mathcal{H} \rightarrow \mathcal{H}$ can be specified completely by describing its action on a basis set of $\mathcal{H}$. To see this let us chose an orthonormal basis $\{|e_i\rangle | i = 1, \ldots, N\}$. Then we can calculate the action of $\hat{A}$ on this basis. We find that the basis $\{|e_i\rangle | i = 1, \ldots, N\}$ is mapped into a new set of vectors $\{|f_i\rangle | i = 1, \ldots, N\}$ following

$$|f_i\rangle := \hat{A}|e_i\rangle \tag{1.53}$$

Of course every vector $|f_i\rangle$ can be represented as a linear combination of the basis vectors $\{|e_i\rangle | i = 1, \ldots, N\}$, i.e.

$$|f_i\rangle := \sum_k A_{ki}|e_k\rangle \tag{1.54}$$

25

Combining Eqs. (1.53) and (1.54) and taking the scalar product with $|e_j\rangle$ we find

$$A_{ji} = \langle e_j| \sum_k \hat{A}_{ki}\, e_k\rangle) \tag{1.55}$$

$$= \langle e_j|f_i\rangle$$

$$= \langle e_j|\hat{A}|e_i\rangle \tag{1.56}$$

The A_{ji} are called the matrix elements of the linear operator $\hat{A}$ with respect to the orthonormal basis $\{|e_i\rangle|i = 1,\dots,N\}$.

We will now go ahead and express linear operators in the Dirac notation. First we will present a particularly simple operator, namely the unit operator $\hat{\mathbb{1}}$, which maps every vector into itself. Surprisingly enough this operator, expressed in the Dirac notation will prove to be very useful in calculations. To find its representation in the Dirac notation, we consider an arbitrary vector $|f\rangle$ and express it in an orthonormal basis $\{|e_i\rangle|i = 1,\dots,N\}$. We find

$$|f\rangle = \sum_{j=1}^{N} f_j\,|e_j\rangle = \sum_{j=1}^{N} |e_j\rangle\langle e_j|f\rangle \tag{1.57}$$

To check that this is correct you just form the scalar product between $|f\rangle$ and any of the basis vectors $|e_i\rangle$. Now let us rewrite Eq. (1.57) a little bit, thereby defining the Dirac notation of operators.

$$|f\rangle = \sum_{j=1}^{N} |e_j\rangle\langle e_j|f\rangle =: \left(\sum_{j=1}^{N} |e_j\rangle\langle e_j|\right)|f\rangle \tag{1.58}$$

The right hand side is defined in terms of the left hand side. The object in the brackets is quite obviously the identity operator because it maps any vector $|f\rangle$ into the same vector $|f\rangle$. Therefore it is totally justified to say that

$$\mathbb{1} \equiv \sum_{j=1}^{N} |e_j\rangle\langle e_j| \tag{1.59}$$

This was quite easy. We just moved some brackets around and we found a way to represent the unit operator using the Dirac notation. Now we can already guess how the general operator will look like, but we will carefully derive it using the identity operator. Clearly we have the following identity

$$\hat{A} = \hat{\mathbb{I}}\hat{A}\hat{\mathbb{I}} \tag{1.60}$$

Now let us use the Dirac notation of the identity operator in Eq. (1.59) and insert it into Eq. (1.60). We then find

$$\hat{A} = \left(\sum_{j=1}^{N}|e_j\rangle\langle e_j|\right)\hat{A}\left(\sum_{k=1}^{N}|e_k\rangle\langle e_k|\right)$$

$$= \left(\sum_{jk}^{N}|e_j\rangle\right)(\langle e_j|\hat{A}\langle e_k|)\langle e_k|$$

$$= \sum_{jk}^{N}(\langle e_j|\hat{A}\langle e_k|)\,|e_j\rangle\langle e_k|$$

$$= \sum_{jk}^{N} A_{jk}|e_j\rangle\langle e_k| \tag{1.61}$$

Therefore you can express any linear operator in the Dirac notation, once you know its matrix elements in an orthonormal basis.

Matrix elements are quite useful when we want to write down linear operator in matrix form. Given an orthonormal basis $\{|e_i\rangle|i = 1,\ldots,N\}$ we can write every vector as a column of numbers

$$|g\rangle = \sum_i g_i|e_i\rangle = \begin{pmatrix} g_1 \\ \vdots \\ g_N \end{pmatrix} \tag{1.62}$$

Then we can write our linear operator in the same basis as a matrix

$$\hat{A} = \begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{pmatrix} \tag{1.63}$$

(1.2.2): Adjoint and Hermitean Operators

Operators that appear in quantum mechanics are linear. But not all linear operators correspond to quantities that can be measured. In this subsection we will describe these operators. In the following subsection we will then discuss some of their properties which then explain why these operators describe measurable quantities.

In the previous section we have considered the Dirac notation and in particular we have seen how to write the scalar product and matrix elements of an operator in this notation. Let us reconsider the matrix elements of an operator $\hat{A}$ in an orthonormal basis $\{|e_i\rangle | i = 1, \ldots, N\}$. We have

$$(\langle e_i | \hat{A} \langle e_j|) = (\langle e_k | \hat{A}) \langle e_j| \tag{1.64}$$

Where we have written the scalar product in two ways. While the left hand side is clear there is now the question, what the bra $\langle e_i | \hat{A}$ on the right hand side means, or better, to which ket it corresponds to. To see this we need to make the,

Definition (1.2.2) :

The adjoint operator $\widehat{A}^\dagger$ corresponding to the linear operator $\widehat{A}$ is the operator such that for all $|x\rangle$, $|y\rangle$ we have

$$(\widehat{A}^\dagger |x\rangle, |y\rangle) := (|x\rangle, \widehat{A}|y\rangle) \tag{1.65}$$

or using the complex conjugation in Eq. (1.65) we have

$$\langle y | \widehat{A}^\dagger | x\rangle := \langle x | \widehat{A} | y\rangle^* \tag{1.66}$$

In matrix notation, we obtain the matrix representing $\hat{A}^\dagger$ by transposition and complex conjugation of the matrix representing $\hat{A}$.

Example (1.2.3) :

$$\hat{A} = \begin{pmatrix} 1 & 2i \\ i & 2 \end{pmatrix} \tag{1.67}$$

and

$$\hat{A}^\dagger = \begin{pmatrix} 1 & -i \\ -2i & 2 \end{pmatrix} \tag{1.68}$$

The following property of adjoint operators is often used.

Lemma (1.2.4) :

For operators $\hat{A}$ and $\hat{B}$ we find

$$\left(\hat{A}\hat{B}\right)^\dagger = \hat{B}^\dagger \hat{A}^\dagger \tag{1.69}$$

Proof

Eq. (1.69) is proven by

$$(|x\rangle, \left(\hat{A}\hat{B}\right)^\dagger |y\rangle = ((\hat{A}\hat{B})|x\rangle, |y\rangle)$$

$$= (\hat{A}(\hat{B}|x\rangle), |y\rangle) \text{ Now use Def. (1.2.2).}$$

$$= (\hat{B}|x\rangle), \hat{A}^\dagger |y\rangle) \text{ Use Def. (1.2.2).again.}$$

$$= (|x\rangle), \hat{B}^\dagger \hat{A}^\dagger |y\rangle)$$

As this is true for any two vectors $|x\rangle$ and $|y\rangle$ the two operators $\left(\hat{A}\hat{B}\right)^\dagger$ and $\hat{B}^\dagger \hat{A}^\dagger$ are equal.

It is quite obvious that in general an operator $\hat{A}$ and its adjoint operator $\hat{A}^\dagger$ are different. However, there are exceptions and these exceptions are very important.

Definition (1.2.5) :

An operator $\hat{A}$ is called Hermitean or self-adjoint if it is equal to its adjoint operator, i.e. if for all states $|x\rangle$, $|y\rangle$ we have

$$\langle y|\hat{A}|x\rangle = \langle x|\hat{A}|y\rangle^* \tag{1.70}$$

In the finite dimensional case a self-adjoint operator is the same as a Hermitean operator. In the infinite-dimensional case Hermitean and self-adjoint are not equivalent.

The difference between self-adjoint and Hermitean is related to the domain of definition of the operators $\hat{A}$ and $\hat{A}^{\dagger}$ which need not be the same in the infinite dimensional case. In the following I will basically always deal with finite-dimensional systems and we will therefore usually use the term Hermitean for an operator that is self-adjoint.

(1.2.3):Eigenvectors, Eigenvalues and the Spectral Theorem

Hermitean operators have a lot of nice properties and we will explore some of these properties. Mostly these properties are concerned with the eigenvalues and eigenvectors of the operators. We start with the definition of eigenvalues and eigenvectors of a linear operator.

Definition (1.2.6) :

A linear operator $\hat{A}$ on an N-dimensional Hilbert space is said to have an eigenvector $|\lambda\rangle$ with corresponding eigenvalue λ if

$$\hat{A}|\lambda\rangle = \lambda|\lambda\rangle \qquad (1.71)$$

or equivalently

$$(\hat{A} - \lambda\mathbb{1})|\lambda\rangle = 0 \qquad (1.72)$$

This definition of eigenvectors and eigenvalues immediately shows us how to determine the eigenvectors of an operator.

Because Eq. (1.72) implies that the N columns of the operator $\hat{A} - \lambda\mathbb{1}$ are linearly dependent we need to have that

$$\det(\hat{A} - \lambda\mathbb{1}) = 0 \qquad (1.73)$$

This immediately gives us a complex polynomial of degree N. As we know from analysis, every polynomial of degree N has exactly N solutions if one includes the multiplicities of the eigenvalues in the counting. In general eigenvalues and eigenvectors do not have many restriction for an arbitrary $\hat{A}$. However, for Hermitean and unitary operators there are a number of nice results concerning the eigenvalues and eigenvectors.

We begin with an analysis of Hermitean operators.

Lemma (1.2.7) :

For any Hermitean operator $\hat{A}$ we have

(1)   All eigenvalues of $\hat{A}$ are real.
(2)   Eigenvectors to different eigenvalues are orthogonal.

Proof

(1)   Given an eigenvalue λ and the corresponding eigenvector $|\lambda\rangle$ of a Hermitean operator $\hat{A}$. Then we have using the hermiticity of $\hat{A}$

$$\lambda^* = \langle\lambda|\hat{A}|\lambda\rangle^* = \langle\lambda|\hat{A}^\dagger|\lambda\rangle = \langle\lambda|\hat{A}|\lambda\rangle = \lambda \tag{1.74}$$

which directly implies that λ is real.

(2) Given two eigenvectors $|\lambda\rangle$, $|\mu\rangle$ for different eigenvalues λ and μ. Then we have

$$\lambda\langle\lambda|\mu\rangle = (\lambda\langle\mu|\lambda\rangle)^* = \left(\langle\mu|\hat{A}|\lambda\rangle\right)^* = \langle\lambda|\hat{A}|\mu\rangle = \mu\langle\lambda|\mu\rangle \tag{1.75}$$

As λ and μ are different this implies $\langle\lambda|\mu\rangle = 0$. This finishes the proof.

Completeness Theorem:

For any Hermitean operator $\hat{A}$ on a Hilbert space $\mathcal{H}$ the set of all eigenvectors form an orthonormal basis of the Hilbert space $\mathcal{H}$, i.e. given the eigenvalues λ_i and the eigenvectors $|\lambda_i\rangle$ we find

$$\hat{A} = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i| \tag{1.76}$$

And for any vector $|x\rangle \in \mathcal{H}$ we find coefficients x_i such that

$$|x\rangle = \sum_i x_i |\lambda_i\rangle \tag{1.77}$$

Now let us briefly consider the case for degenerate eigenvalues. This is the case, when the characteristic polynomial has multiple zero's. In other words, an eigenvalue is degenerate if there is more than one eigenvector corresponding to it. An eigenvalue λ is said to be fold degenerate if there is a set of d linearly

independent eigenvectors $\{|\lambda_1\rangle, \ldots, |\lambda_d\rangle\}$ all having the same eigenvalue λ. Quite obviously the space of all linear combinations of the vectors $\{|\lambda_1\rangle, \ldots, |\lambda_d\rangle\}$ is a d-dimensional vector space. Therefore we can find an orthonormal basis of this vector space. This implies that any Hermitean operator has eigenvalues $\lambda_1, \ldots, \lambda_k$ with degeneracies $d(\lambda_1), \ldots, d(\lambda_k)$. To each eigenvector λ_i we can find an orthonormal basis of $d(\lambda_i)$ vectors. Therefore the above completeness theorem remains true also for Hermitean operators with degenerate eigenvalues.

Now you might wonder whether every linear operator $\hat{A}$ on an N dimensional Hilbert space has N linearly independent eigenvectors? It turns out that this is not true. An example is the 2×2 matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

which has only one eigenvalue $\lambda = 0$. Therefore any eigenvector to this eigenvalue has to satisfy

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which implies that $b = 0$. But then the only normalized eigenvector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and therefore the set of eigenvectors do not form an orthonormal basis.

We have seen that any Hermitean operator $\hat{A}$ can be expanded in its eigenvectors and eigenvalues. The procedure of finding this particular representation of an operator is called diagonalization. Often we do not want to work in the basis of the eigenvectors but for example in the canonical basis of the vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_N = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{1.78}$$

If we want to rewrite the operator $\hat{A}$ in that basis we need to find a map between the canonical basis and the orthogonal basis of the eigenvectors of $\hat{A}$. If we write

the eigenvectors $|\lambda_i\rangle = \sum_{j=1}^{N} \alpha_{ji} |e_j\rangle$ then this map is given by the unitary operator (we will define unitary operators shortly)

$$\hat{U} = \sum_{i=1}^{N} |\lambda_j\rangle \langle e_i| = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1N} \\ \vdots & \ddots & \vdots \\ \alpha_{N1} & \cdots & \alpha_{NN} \end{pmatrix} \tag{1.79}$$

Which obviously maps a vector $|e_i\rangle$ into the eigenvector corresponding to the eigenvalue $|\lambda_i\rangle$. Using this operator $\hat{U}$ we find

$$\hat{U}^{\dagger} \hat{A} \, \hat{U} = \sum_{i} \lambda_i \, \hat{U}^{\dagger} |\lambda_i\rangle\langle\lambda_i| \hat{U} = \sum_{i} \lambda_i \, |e_i\rangle\langle e_i| \tag{1.80}$$

The operator in Eq. (1.79) maps orthogonal vectors into orthogonal vectors. In fact, it preserves the scalar product between any two vectors.

Let us use this as the defining property of a unitary transformation.

Definition (1.2.8) :

A linear operator $\hat{U}$ on a Hilbert space $\mathcal{H}$ is called unitary if it is defined for all vectors $|x\rangle$, $|y\rangle$ in $\mathcal{H}$, maps the whole Hilbert space into the whole Hilbert space and satisfies

$$\langle x|\hat{U}^{\dagger}\hat{U}|y\rangle = \langle x|y\rangle \tag{1.81}$$

In fact we can replace the last condition by demanding that the operator satisfies

$$\hat{U}^{\dagger}\hat{U} = \mathbb{1} \quad \text{and} \quad \hat{U}\hat{U}^{\dagger} = \mathbb{1} \tag{1.82}$$

Eq. (1.81) implies that a unitary operator preserves the scalar product and therefore in particular the norm of vectors as well as the angle between any two vectors.

Theorem (1.2.9) :

Any unitary operator $\hat{U}$ on an N-dimensional Hilbert space $\mathcal{H}$ has a complete basis of eigenvectors and all the eigenvalues are of the form $e^{i\phi}$ with real ϕ.

Proof

We will not give a proof that the eigenvectors of $\hat{U}$ form a basis in $\mathcal{H}$. What we will proof is that the eigenvalues of $\mathcal{H}$ are of the form $e^{i\phi}$ with real ϕ. To see this, we use Eq. (1.82). Be $|\lambda\rangle$ an eigenvector of $\hat{U}$ to the eigenvalue λ, then

$$\lambda \hat{U}^{\dagger}|\lambda\rangle = \hat{U}^{\dagger}\hat{U}|\lambda\rangle = |\lambda\rangle \tag{1.83}$$

This implies that $\lambda \neq 0$ because otherwise the right-hand side would be the null-vector, which is never an eigenvector. From Eq. (1.83) we find

$$\frac{1}{\lambda} = \langle\lambda|\hat{U}^{\dagger}|\lambda\rangle = \langle\lambda|\hat{U}|\lambda\rangle^* = \lambda^* \tag{1.84}$$

This results in

$$|\lambda|^2 = 1 \Longleftrightarrow \lambda = e^{i\phi} \tag{1.85}$$

(1.2.4): Functions of Operators

Definition (1.2.10) :

Given an operator $\hat{A}$ with eigenvalues a_i and a complete set of eigenvectors $|a_i\rangle$. Further have a function $f : C \to C$ that maps complex numbers into complex numbers then we define

$$f(\hat{A}) = \sum_{i=1}^{N} f(a_i)|a_i\rangle\langle a_i| \tag{1.86}$$

Definition (1.2.11) :

Given a function $f : C \to C$ that can be expanded into a power series

$$f(z) = \sum_{i=1}^{\infty} f_i \, z^i \tag{1.87}$$

Then we define

$$f(\hat{A}) = \sum_{i=1}^{\infty} f_i \, A^i \tag{1.88}$$

Definition (1.2.12) :

The derivative of an operator function $f(\hat{A})$ is defined via $g(z) = \frac{df}{dz}(z)$ as

$$\frac{df(\hat{A})}{d\hat{A}}(z) = g(\hat{A}) \tag{1.89}$$

Let us see whether the two definitions Def. (1.2.10) and (1.2.11) coincide for operators with complete set of eigenvectors and functions that can be expanded into a power series given in Eq. (1.87).

$$f(\hat{A}) = \sum_{k=1}^{\infty} f_k A^k$$

$$= \sum_{k=1}^{\infty} f_k \left(\sum_{j=1}^{N} a_j |a_j\rangle\langle a_j| \right)^k$$

$$= \sum_{k=1}^{\infty} f_k \sum_{j=1}^{N} a_j^k |a_j\rangle\langle a_j|$$

$$= \sum_{j=1}^{N} \left(\sum_{k=1}^{\infty} f_k a_j^k \right) |a_j\rangle\langle a_j|$$

$$= \sum_{j=1}^{N} f(a_j) |a_j\rangle\langle a_j| \tag{1.90}$$

For operators that do not have a complete orthonormal basis of eigenvectors of eigenvectors it is not possible to use Definition (1.2.10) and we have to go back to Definition (1.2.11). In practise this is not really a problem in quantum mechanics because we will always encounter operators that have a complete set of eigenvectors.

As an example consider a Hermitean operator $\hat{A}$ with eigenvalues a_k and eigenvectors $|a_k\rangle$ and compute $\hat{U} = e^{i\hat{A}}$. We find

$$\hat{U} = e^{i\hat{A}} = \sum_{k=1}^{N} e^{ia_k} |a_k\rangle\langle a_k| \tag{1.91}$$

This is an operator which has eigenvalues of the form e^{ia_k} with real a_k. Therefore it is a unitary operator, which you can also check directly from the requirement $\hat{U}\hat{U}^\dagger = \mathbb{1} = \hat{U}^\dagger\hat{U}$. In fact it is possible to show that every unitary operator can be written in the form Eq. (1.91).

Lemma (1.2.13) :

To any unitary operator $\hat{U}$ there is a Hermitean operator $\hat{H}$ such that

$$\hat{U} = e^{i\hat{H}} \tag{1.92}$$

Show that for any unitary operator $\hat{U}$ we have $f(\hat{U}^\dagger\hat{A}\hat{U}) = \hat{U}^\dagger f(\hat{A})\hat{U}$

Proof

We use the fact that $\hat{U}\hat{U}^\dagger = \mathbb{1}$ to find

$$f(\hat{U}^\dagger\hat{A}\hat{U}) = \sum_{k=0}^{\infty} f_k (\hat{U}^\dagger\hat{A}\hat{U})^k$$

$$= \sum_{k=0}^{\infty} f_k \hat{U}^\dagger\hat{A}^k\hat{U}$$

$$= \hat{U}^\dagger \left(\sum_{k=0}^{\infty} f_k \hat{A}^k \right) \hat{U}$$

$$= \hat{U}^\dagger f(\hat{A})\hat{U}$$

If you have functions, then you will also expect to encounter derivatives of functions. Therefore we have to consider how to take derivatives of matrices. To take a derivative we need to have not only one operator, but a family of operators that is parametrized by a real parameter s. An example is the set of operators of the form

$$\hat{A}(s) = \begin{pmatrix} 1 + s & i.s \\ -i.s & 1 - s \end{pmatrix} \tag{1.93}$$

Another example which is familiar to you is the time evolution operator $e^{-i\hat{H}s}$.

Now we can define the derivative with respect to s in complete analogy to the derivative of a scalar function by

$$\frac{d\hat{A}}{ds}(s) := \lim_{\Delta s \to 0} \frac{\hat{A}(s + \Delta s) - \hat{A}(s)}{\Delta s} \tag{1.94}$$

This means that we have defined the derivative of an operator component wise.

Now let us explore the properties of the derivative of operators. First let us see what the derivative of the product of two operators is. We find

Property (1.2.14):

For any two linear operators $\hat{A}(s)$ and $\hat{B}(s)$ we have

$$\frac{d(\hat{A}\hat{B})}{ds}(s) = \frac{d\hat{A}}{ds}(s)\hat{B}(s) + \hat{A}(s)\frac{d\hat{B}}{ds}(s) \tag{1.95}$$

This looks quite a lot like the product rule for ordinary functions, except that now the order of the operators plays a crucial role.

You can also have functions of operators that depend on more than one variables. A very important example is the commutator of two operators.

Definition (1.2.15):

For two operators $\hat{A}$ and $\hat{B}$ the commutator $[\hat{A}, \hat{B}]$ is defined as

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} \tag{1.96}$$

While the commutator between numbers (1×1 matrices) is always zero, this is not the case for general matrices. For example the operators corresponding to momentum and position do not commute, i.e. their commutator is nonzero. Other examples are the Pauli spin-operators

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.97}$$

For $i, j = 1, 2, 3$ they have the commutation relations

$$[\sigma_i, \sigma_j] = i\epsilon_{ijk}\sigma_k \tag{1.98}$$

Where ϵ_{ijk} is the completely antisymmetric tensor. It is defined by $\epsilon_{123} = 1$ and changes sign, when two indices are interchanged, for example $\epsilon_{ijk} = -\epsilon_{jik}$.

Lemma (1.2.16):

For arbitrary linear operators $\hat{A}, \hat{B}, \hat{C}$ on the same Hilbert space we have

$$[\hat{A}\hat{B}, \hat{C}] = \hat{A}[\hat{B}, \hat{C}] + [\hat{A}, \hat{C}]\hat{B} \tag{1.99}$$

$$0 = [\hat{A}, [\hat{B}, \hat{C}]] + [\hat{B}, [\hat{C}, \hat{A}]] + [\hat{C}, [\hat{A}, \hat{B}]] \tag{1.100}$$

Lemma (1.2.17):

Two commuting observables $\hat{A}$ and $\hat{B}$ have the same eigenvectors, i.e. they can be diagonalized simultaneously.

Proof

For simplicity we assume that both observables have only nondegenerate eigenvalues.

Now chose a basis of eigenvectors $\{|a_i\rangle\}$ that diagonalizes $\hat{A}$. Now try to see whether the $|a_i\rangle$ are also eigenvectors of $\hat{B}$. Using $[\hat{A}, \hat{B}]$ we have

$$\hat{A}(\hat{B}|a_i\rangle) = \hat{B}\hat{A}|a_i\rangle = a_i(\hat{B}|a_i\rangle) \tag{1.101}$$

This implies that $\hat{B}|a_i\rangle$ is an eigenvector of $\hat{A}$ with eigenvalue a_i. As the eigenvalue is non-degenerate we must have

$$\hat{B}|a_i\rangle = b_i|a_i\rangle \tag{1.102}$$

For some bi. Therefore $|a_i\rangle$ is an eigenvector to $\hat{B}$.

Lemma (1.2.18) :

38

Given two linear operators $\hat{A}$ and $\hat{B}$ which have the commutator $[\hat{B}, \hat{A}] = \mathbb{1}$. Then for the derivative of an operator function $f(\hat{A})$ we find

$$[\hat{B}, f(\hat{A})] = \frac{df}{d\hat{A}}(\hat{A}) \tag{1.103}$$

Proof

Remember that a function of an operator is defined via its expansion into a power series, see Eq. (1.88). Therefore we find

$$[\hat{B}, f(\hat{A})] = \left[\hat{B}, \sum_k f_k \hat{A}^k\right]$$

$$= \sum_k f_k [\hat{B}, \hat{A}^k]$$

Now we need to evaluate the expression $[\hat{B}, \hat{A}^k]$. We proof by induction that $[\hat{B}, \hat{A}^n] = n\hat{A}^{n-1}$. For $n = 1$, this is true. Assume that the assumption is true for n. Now start with $n + 1$ and reduce it to the case for n. Using Eq. (1.99) we find

$$[\hat{B}, \hat{A}^{(n+1)}] \equiv [\hat{B}, \hat{A}^n \hat{A}] = [\hat{B}, \hat{A}^n]\hat{A} + \hat{A}^n[\hat{B}, \hat{A}] \tag{1.104}$$

Now using the $[\hat{B}, \hat{A}] = \mathbb{1}$ and the induction assumption we find

$$[\hat{B}, \hat{A}^{(n+1)}] = n\hat{A}^{n-1}\hat{A} + \hat{A}^n = (n + 1)\hat{A}^n \tag{1.105}$$

Now we can conclude

$$[\hat{B}, f(\hat{A})] = \sum_k f_k [\hat{B}, \hat{A}^k]$$

$$= \sum_k f_k k \, \hat{A}^{k-1}$$

$$= \frac{df}{d\hat{A}}(\hat{A})$$

This finishes proof.

A very useful property is

Lemma (1.2.19) :

(Baker-Campbell-Haussdorff) For general operators $\hat{A}$ and $\hat{B}$ we have

$$e^{\hat{B}}\hat{A}e^{-\hat{B}} = \hat{A} + \left[\hat{B},\hat{A}\right] + \frac{1}{2}\left[\hat{B},\left[\hat{B},\hat{A}\right]\right] + \cdots \qquad (1.106)$$

For operators such that $\left[\hat{B},\left[\hat{B},\hat{A}\right]\right] = 0$ we have the simpler version

$$e^{\hat{B}}\hat{A}e^{-\hat{B}} = \hat{A} + \left[\hat{B},\hat{A}\right] \qquad (1.107)$$

Proof

Define the function of one real parameter α

$$f(\alpha) = e^{\alpha\hat{B}}\hat{A}e^{-\alpha\hat{B}} \qquad (1.108)$$

We can expand this function around $\alpha = 0$ into a Taylor series $f(\alpha) = \sum_{n=0}^{\infty}\frac{\alpha^n}{n!}\frac{d^n f}{d\alpha^n}(\alpha)|_{\alpha=0}$ and therefore we need to determine the derivatives of the function $f(\alpha)$. We find

$$\frac{df}{d\alpha}(\alpha)|_{\alpha=0} = \left[\hat{B},\hat{A}\right]$$

$$\frac{d^2 f}{d\alpha^2}(\alpha)|_{\alpha=0} = \left[\hat{B},\left[\hat{B},\hat{A}\right]\right]$$

$$\vdots$$

$$\vdots$$

The rest of the proof follows by induction. The proof of Eq. (1.107) follows directly from Eq. (1.106).

(1.3): Operators with Continuous Spectrum

(1.3.1):The Position Operator

The most natural place where an infinite dimensional state space appears is in the analysis of a particle in free space. Therefore let us briefly reconsider some aspects of wave mechanics. The state of a particle in free space (maybe moving in a potential) is described by the square-integrable wave-function $\psi(x)$.

The question is now as to how we connect the wave-function notation with the Dirac notation which we have used to develop our theory of Hilbert spaces and linear operators.

Let us remember what the Dirac notation for finite dimensional systems means mathematically. Given the ket-vector $|\phi\rangle$ for the state of a finite dimensional system, we can find the components of this ketvector with respect to a certain basis $\{|e_i\rangle\}$. The i-th component is given by the complex number $\langle e_i|\phi\rangle$.

Therefore in a particular basis it makes sense to write the state $|\phi\rangle$ as a column vector

$$|\phi\rangle \leftrightarrow \begin{pmatrix} \langle e_1|\phi\rangle \\ \vdots \\ \langle e_n|\phi\rangle \end{pmatrix} \tag{1.109}$$

Let us try to transfer this idea to the wave-function of a particle in free space. What we will do is to interpret $\psi(x)$ as the component of a vector with infinitely many components. Informally written this means

$$|\psi\rangle \leftrightarrow \begin{pmatrix} \vdots \\ \psi(x) \\ \vdots \end{pmatrix} \tag{1.110}$$

where we have given the column vector a name, $|\psi\rangle$. Obviously the set of vectors defined in this way form a vector space as you can easily check. Of course we would like to have a scalar product on this vector space. This is introduced in complete analogy to the finite dimensional case. There we had

$$(|\phi\rangle, |\psi\rangle) = \sum_i \langle\phi|e_i\rangle \langle e_i|\psi\rangle \tag{1.111}$$

We just replace the summation over products of components of the two vectors by an integration. We have (see also Eq (1.29)

$$(|\phi\rangle, |\psi\rangle) := \int_{-\infty}^{\infty} dx\, \phi^*(x)\psi(x) \tag{1.112}$$

Now we have a space of vectors (or square integrable wave functions) $\mathcal{H}$ equipped with a scalar product. Indeed it turns out to be a complete space so that we have a Hilbert space.

Now that we have introduced ket vectors we also need to define bra-vectors. In the finite dimensional case we obtained the bra vector via the idea of linear functionals on the space of state vectors. Let us repeat this procedure now for the the case of infinite dimensions. We define a linear functional $\langle\phi|$ by

$$\langle\phi|(|\psi\rangle) \equiv \langle\phi|\psi\rangle = (|\phi\rangle, |\psi\rangle) \tag{1.113}$$

Now we would like to investigate a particular ket vector (linear functional) that will allow as to define a position state. We define a linear functional $\langle x_0|$ by

$$\langle x_0|(|\psi\rangle) \equiv \langle x_0|\psi\rangle := \psi(x_0) \tag{1.114}$$

We are already writing this functional very suggestively as a bra-vector. Thats perfectly ok, and we just have to check that the so defined functional is indeed linear. Of course we would like to interpret the left hand side of Eq. (1.114) as a scalar product between two ket's, i.e.

$$(|x_0\rangle, |\psi\rangle) := \langle x_0|\psi\rangle = \psi(x_0) \tag{1.115}$$

Using the scalar product Eq. (1.112), we have

$$\int_{-\infty}^{\infty} dx \, \delta_{x_0}^*(x)\psi(x) = (|x_0\rangle, |\psi\rangle) = \langle x_0|\psi\rangle = \psi(x_0) \tag{1.116}$$

This means that the function $\delta_{x_0}^*(x)$ has to act like a delta-function! The wave-function corresponding to the bra $\langle x_0|$ is a delta-function. A delta-function however, is not square-integrable! Therefore it cannot be an element of the Hilbert space of square integrable functions. However, as we have seen it would be quite convenient to use these wave-functions or states. Therefore we just add them to our Hilbert space, although we will often call them improper states or wave-functions. In fact we can use basically all the rules that we have learned about finite dimensional Hilbert-spaces also for these improper states. All we need to demand is the following rule for the scalar product

$$\langle\psi|x_0\rangle := \langle x_0|\psi\rangle^* = \psi^*(x_0) \tag{1.117}$$

42

Now we can write for arbitrary kets $|\phi\rangle, |\psi\rangle \in \mathcal{H}$

$$\langle\phi|\psi\rangle = \int \phi^*(x)\,\psi(x)dx = \int \langle\phi|x\rangle\,\langle x|\psi\rangle dx = \langle\phi| \left(\int |x\rangle\langle x|\,dx\right) |\psi\rangle \quad (1.118)$$

Then we can conclude

$$\int |x\rangle\langle x|\,dx = 1 \qquad (1.119)$$

Inserting this identity operator in $\langle x|\psi\rangle$, we obtain the orthogonality relation between position kets

$$\int \delta(x-x')\,\psi(x')dx' = \psi(x) = \langle x|\psi\rangle = \int \langle x|x'\rangle\,\langle x'|\psi\rangle dx'$$

$$= \int \langle x|x'\rangle\,\psi(x')dx'$$

Therefore we have

$$\langle x|x'\rangle = \delta(x-x') \qquad (1.120)$$

Now we can derive the form of the position operator from our knowledge of the definition of the position expectation value

$$\langle\psi|\hat{x}|\psi\rangle := \int x\,|\psi(x)|^2 dx$$

$$= \int \langle\psi|x\rangle\,x\langle x|\psi\rangle dx$$

$$= \langle\psi| \left(\int x|x\rangle\,dx\right) |\psi\rangle \qquad (1.121)$$

where we defined the position operator

$$\hat{x} = \int x|x\rangle\langle x|\,dx = \hat{x}^\dagger \qquad (1.122)$$

Now we see why the improper position kets are so useful. In this basis the position operator is automatically diagonal. The improper position kets $|x_0\rangle$ are eigenvectors of the position operator

$$\hat{x}|x_0\rangle = x_0|x_0\rangle \tag{1.123}$$

This makes sense, as the position kets describe a particle that is perfectly localized at position x_0. Therefore a position measurement should always give the result x_0. So far we have dealt with one-dimensional systems. All of the above considerations can be generalized to the d-dimensional case by setting

$$|\hat{x}\rangle = \hat{x}_1\vec{e}_1 + \cdots + \hat{x}_d\vec{e}_d \tag{1.124}$$

The different components of the position operator commute are assumed to commute.

(1.3.2):The Momentum Operator

Now we are going to introduce the momentum operator and momentum eigenstates using the ideas of linear functionals in a similar fashion to the way in which we introduced the position operator. Let us introduce the linear functional $\langle p|$ defined by

$$\langle p|\psi\rangle := \frac{1}{\sqrt{2\pi h}}\int e^{-ipx/h}\,\psi(x)dx \tag{1.125}$$

Now we define the corresponding ket by

$$\langle p|\psi\rangle^* = \hat{\psi}^*(p) =: \langle\psi|p\rangle \tag{1.126}$$

Combining Eq. (1.125) with the identity operator as represented in Eq. (1.119) we find

$$\frac{1}{\sqrt{2\pi h}}\int e^{-ipx/h}\,\psi(x)dx = \langle p|\psi\rangle = \int\langle p|x\rangle\,\langle x|\psi\rangle dx \tag{1.127}$$

Therefore we find that the state vector $|p\rangle$ represents a plane wave, because

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi h}}e^{ipx/h} \tag{1.128}$$

As Eq. (1.128) represents a plane wave with momentum p it makes sense to call $|p\rangle$ a momentum state and expect that it is an eigenvector of momentum operator $\hat{p}$ to the eigenvalue p. Before we define the momentum operator, let us

find the decomposition of the identity operator using momentum eigenstates. To see this we need to remember from

the theory of delta-functions that

$$\frac{1}{\sqrt{2\pi h}}\int e^{ip(x-y)/h}\,dp = \delta(x-y) \tag{1.129}$$

Then we have for arbitrary $|x\rangle$ and $|y\rangle$

$$\langle x|y\rangle = \delta(x-y) = \frac{1}{\sqrt{2\pi h}}\int e^{ip(x-y)/h} = \langle x|\left(\int|p\rangle\langle p|\,dp\right)|y\rangle \tag{1.130}$$

and therefore

$$\int|p\rangle\langle p|\,dp = \mathbb{1} \tag{1.131}$$

The orthogonality relation between different momentum kets can be found by using Eq. (1.131) in Eq. (1.125).

$$\langle p|\psi\rangle = \langle p|\mathbb{1}|\psi\rangle = \int\langle p|p'\rangle\,\langle p'|\psi\rangle dp' \tag{1.132}$$

So that

$$\langle p|p'\rangle = \delta(p-p') \tag{1.133}$$

The momentum operator $\hat{p}$ is the operator that has as its eigenvectors the momentum eigenstates $|p\rangle$ with the corresponding eigenvalue p. This makes sense, because $|p\rangle$ describes a plane wave which has a perfectly defined momentum. Therefore we know the spectral decomposition which is

$$\hat{p} = \int p|p\rangle\langle p|\,dp \tag{1.134}$$

Clearly we have

$$\hat{p}|p_0\rangle = p_0|p_0\rangle \tag{1.135}$$

Analogously to the position operator we can extend the momentum operator to the d-dimensional space by

$$\hat{\vec{p}} = \hat{p}_1 \vec{e}_1 + \cdots + \hat{p}_d \vec{e}_d \tag{1.136}$$

The different components of the momentum operator are assumed to commute.

(1.3.3): The Position Representation of the Momentum Operator and the Commutator Between Position and Momentum

We have seen how to express the position operator in the basis of the improper position kets and the momentum operator in the basis of the improper momentum kets. Now we would like to see how the momentum operator looks like in the position basis.

To see this, differentiate Eq. (1.128) with respect to χ which gives

$$\frac{h}{i} \frac{\partial}{\partial x} \langle x|p \rangle = \frac{h}{i} \frac{\partial}{\partial x} \frac{1}{\sqrt{2\pi h}} e^{ipx/h} = p\langle x|p \rangle \tag{1.137}$$

Therefore we find

$$\langle x|\hat{p}|\psi \rangle = \int \langle x|p \rangle \, p \langle p|\psi \rangle \, dp$$

$$= \frac{h}{i} \frac{\partial}{\partial x} \int \langle x|p \rangle \langle p|\psi \rangle \, dp$$

$$= \frac{h}{i} \langle x| \left(\int |p \rangle \langle p| \right) |\psi \rangle$$

$$= \frac{h}{i} \frac{\partial}{\partial x} \langle x|\psi \rangle \tag{1.138}$$

In position representation the momentum operator acts like the differential operator, i.e.

$$\hat{p} \longleftrightarrow \frac{h}{i} \frac{\partial}{\partial x} \tag{1.139}$$

Knowing this we are now able to derive the commutation relation between momentum and position operator. We find

$$\langle x|[\hat{x},\hat{p}]|\psi \rangle = \langle x|(\hat{x}\hat{p} - \hat{p}\hat{x}|\psi \rangle)$$

$$= \frac{h}{i} \left[x \frac{\partial}{\partial x} \langle x | \psi \rangle - \frac{\partial}{\partial x} (x \langle x | \psi \rangle) \right]$$

$$= ih \langle x | \psi \rangle$$

$$= \langle x | ih \mathbb{1} | \psi \rangle$$

Therefore we have the Heisenberg commutation relations

$$[\hat{x}, \hat{p}] = ih \mathbb{1} \qquad (1.140)$$