Introduction to Quantum Physics

Rainer F. Hauser
rainer.hauser@gmail.com

January 7, 2025

Abstract

The video channel with the title "Professor M does Science" in YouTube offers a simple step-by-step but all the same very valuable and rigorous introduction into the world of quantum mechanics. This script covers the postulates of quantum physics and helps to digest the topic covered by a group of those videos but is not meant as a replacement for them.

1 The Postulates of Quantum Mechanics

1.1 State Space, Dual Space and Dirac Notation

State space is the name for the vector space in which the quantum systems exist. The three-dimensional Euclidean space $\mathbb{R}^3$ is the space for classical physics. State space can have any finite or infinite dimensions, and its vectors are described by complex numbers and not by real numbers. Euclidean spaces and state spaces are called Hilbert spaces and have an *inner product* called scalar product.

Postulate I: The state of a physical system is characterized by a *state vector* that belongs to a complex vector space $\mathcal{V}$ which is called the *state space* of the system.

Elements in the Euclidean space $\mathbb{R}^3$ are called vectors, are written as $\underline{r}$ and satisfy the laws of addition as well as the laws of scalar multiplication with $a \in \mathbb{R}$. Similarly, elements of the state space $\mathcal{V}$ are called "ket", are written as $|\psi\rangle$ in the Dirac notation, and satisfy the laws of addition as well as the laws of multiplication with a scalar $a \in \mathbb{C}$. The laws of vector addition and scalar product are on the left side for Euclidean vectors $\underline{r}$ and on the right side for state vectors $|\psi\rangle$

$\underline{r}_1 + \underline{r}_2 = \underline{r}_3 \in \mathbb{R}^3$	$\lvert\psi_1\rangle + \lvert\psi_2\rangle = \lvert\psi_3\rangle \in \mathcal{V}$
$\underline{r}_1 + \underline{r}_2 = \underline{r}_2 + \underline{r}_1$	$\lvert\psi_1\rangle + \lvert\psi_2\rangle = \lvert\psi_2\rangle + \lvert\psi_1\rangle$
$(\underline{r}_1 + \underline{r}_2) + \underline{r}_3 = \underline{r}_1 + (\underline{r}_2 + \underline{r}_3)$	$(\lvert\psi_1\rangle + \lvert\psi_2\rangle) + \lvert\psi_3\rangle = \lvert\psi_1\rangle + (\lvert\psi_2\rangle + \lvert\psi_3\rangle)$
$\underline{0} + \underline{r} = \underline{r}$	$0 + \lvert\psi\rangle = \lvert\psi\rangle$
$\underline{r} + (-\underline{r}) = \underline{0}$	$\lvert\psi\rangle + (-\lvert\psi\rangle) = 0$
$a\underline{r} \in \mathbb{R}^3$	$a\lvert\psi\rangle \in \mathcal{V}$
$a_1(a_2\underline{r}) = (a_1 a_2)\underline{r}$	$a_1(a_2\lvert\psi\rangle) = (a_1 a_2)\lvert\psi\rangle$
$(a_1 + a_2)\underline{r} = a_1\underline{r} + a_2\underline{r}$	$(a_1 + a_2)\lvert\psi\rangle = a_1\lvert\psi\rangle + a_2\lvert\psi\rangle$
$a(\underline{r}_1 + \underline{r}_2) = a\underline{r}_1 + a\underline{r}_2$	$a(\lvert\psi_1\rangle + \lvert\psi_2\rangle) = a\lvert\psi_1\rangle + a\lvert\psi_2\rangle$
$1\underline{r} = \underline{r}$	$1\lvert\psi\rangle = \lvert\psi\rangle$

with the null vector $\underline{0}$ and the inverse $-\underline{r}$ for the Euclidean space, as well as the null ket 0 and the inverse ket $-|\psi\rangle$ for the state space.

The scalar product $SP(\underline{r}_1, \underline{r}_2) = c \in \mathbb{R}$ in Euclidean space has the properties

1. Conjugation: $SP(\underline{r}_1, \underline{r}_2) = SP(\underline{r}_2, \underline{r}_1)$

2. Linearity:
   $SP(\underline{r}_1, a\underline{r}_2) = a\,SP(\underline{r}_1, \underline{r}_2)$, $SP(a\underline{r}_1, \underline{r}_2) = a\,SP(\underline{r}_1, \underline{r}_2)$
   $SP(\underline{r}_1, \underline{r}_2 + \underline{r}_3) = SP(\underline{r}_1, \underline{r}_2) + SP(\underline{r}_1, \underline{r}_3)$
3. Positivity: $SP(\underline{r}, \underline{r}) \geq 0$ and $SP(\underline{r}, \underline{r}) = 0 \Leftrightarrow \underline{r} = \underline{0}$

where $SP(\underline{r}_1, \underline{r}_2)$ is usually written as $\underline{r}_1 \cdot \underline{r}_2$ with $\underline{r} \cdot \underline{r} = |\underline{r}|^2$. Similarly, the scalar product $SP(|\psi\rangle, |\varphi\rangle) = c$ with $c \in \mathbb{C}$ for the state space $\mathcal{V}$ has the properties

1. Conjugation: $SP(|\psi\rangle, |\varphi\rangle) = [SP(|\varphi\rangle, |\psi\rangle)]^*$
2. Linearity and antilinearity:
   $SP(|\psi\rangle, a\,|\varphi\rangle) = a\,SP(|\psi\rangle, |\varphi\rangle) \Rightarrow$ is linear in the second argument
   $SP(a\,|\psi\rangle, |\varphi\rangle) = [SP(|\varphi\rangle, a\,|\psi\rangle)]^* = a^*[SP(|\varphi\rangle, |\psi\rangle]^* = a^*\,SP(|\psi\rangle, |\varphi\rangle) \Rightarrow$ is antilinear in the first argument
   $SP(|\psi\rangle, |\varphi\rangle + |\chi\rangle) = SP(|\psi\rangle, |\varphi\rangle) + SP(|\psi\rangle, |\chi\rangle)$
3. Positivity: $SP(|\psi\rangle, |\psi\rangle) \geq 0$ and $SP(|\psi\rangle, |\psi\rangle) = 0 \Leftrightarrow |\psi\rangle = 0$

where the differences to the Euclidean space comes from the fact that the scalar product here is a complex number where c^* is the complex conjugate of c.

The scalar product in Euclidean space can be written as

$$SP(\underline{r}_1, \underline{r}_2) = \underline{r}_1 \cdot \underline{r}_2 = (x_1, y_1, z_1) \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = x_1x_2 + y_1y_2 + z_1z_2$$

where the vector $\underline{r}_1$ is a "row" vector and $\underline{r}_2$ is a "column" vector. Since one cannot add a row vector and a column vector, they must belong to different vector spaces. This distinction between row and column vectors is important in the state space. The column vectors belong to the vector space and the row vector belongs to the *dual space*.

In terms of *linear maps* the row vector maps a column vector to a scalar. Thus, $SP(\underline{r}_1, \cdot)$ maps $\underline{r}_2$ to a scalar c. In this language the row vector, and equivalently $SP(\underline{r}_1, \cdot)$, are objects that act on column vectors in the vector space. There is a one-to-one correspondence between column and row vectors, and therefore there is a one-to-one correspondence between vectors and linear maps. It can be shown by checking the defining properties above for a vector space that the set $SP(\underline{r}_1, \cdot)$ of linear maps forms a vector space. Thus, the dual space is in other words the vector space made of the linear maps that act on the original vector space.

Similarly, for the state space $SP(|\psi\rangle, \cdot)$ is an element of the dual space which is written as $\langle\psi|$ and called "bra" in the Dirac terminology. A ket $|\psi\rangle \in \mathcal{V}$ corresponds to a bra $\langle\psi| \in \mathcal{V}^*$ where $\mathcal{V}^*$ is the dual space. This correspondence is written as $|\psi\rangle \leftrightarrow \langle\psi|$ in the following. The scalar product $SP(|\psi\rangle, |\varphi\rangle)$ can be written as a "bra(c)ket" in the form $\langle\psi|\varphi\rangle$. A subtlety comes from the fact that a is complex and that the scalar product is antilinear in the first argument such that $a\,|\psi\rangle$ corresponds to $a^*\,\langle\psi|$.

1.2 Operators in Quantum Mechanics

An *operator* is the mathematical object that allows to describe physical properties in quantum mechanics. The position operator $\hat{x}$, for example, specifies where the particle is in space, the momentum operator $\hat{p}$ tells what the momentum of the particle is, and $\hat{H}$ is the energy operator called Hamiltonian.

Postulate II: A physical quantity $\mathcal{A}$ is described by an *operator* $\hat{A}$ acting on the state space $\mathcal{V}$, and this operator is an *observable*.

This means that operators act on kets as $\hat{A}\,|\psi\rangle = |\psi'\rangle$ where $|\psi\rangle, |\psi'\rangle \in \mathcal{V}$. Operators in quantum mechanics act on superpositions of kets as

$$\hat{A}\big(a_1\,|\psi_1\rangle + a_2\,|\psi_2\rangle\big) = a_1\,\hat{A}\,|\psi_1\rangle + a_2\,\hat{A}\,|\psi_2\rangle$$

where $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{V}$ and $a_1, a_2 \in \mathbb{C}$. They are therefore called linear operators.

Operators have the properties for addition

- Associativity: $(\hat{A}+\hat{B})+\hat{C}=\hat{A}+(\hat{B}+\hat{C})$
- Commutativity: $\hat{A}+\hat{B}=\hat{B}+\hat{A}$

and the property for multiplication

- Associativity: $(\hat{A}\hat{B})\hat{C}=\hat{A}(\hat{B}\hat{C})$

where multiplication is defined as $(\hat{A}\hat{B})\,|\psi\rangle=\hat{A}(\hat{B}\,|\psi\rangle)=\hat{A}\,|\psi'\rangle$ for $|\psi'\rangle=\hat{B}\,|\psi\rangle$. However, multiplication is usually not commutative such that $\hat{A}\hat{B}\neq\hat{B}\hat{A}$, and this leads to the definition of the *commutator*

$$\left[\hat{A},\hat{B}\right]=\hat{A}\hat{B}-\hat{B}\hat{A} \tag{1.1}$$

for two operators $\hat{A}$ and $\hat{B}$. Operators that do not commute are associated with physical properties that cannot be measured simultaneously. An example for operators that do not commute are the position and the momentum operators, and therefore the position and momentum operators of a particle can never be measured simultaneously.

The scalar product of $|\psi\rangle$ with $\hat{A}\,|\varphi\rangle=|\varphi'\rangle$ is $\langle\psi|\varphi'\rangle=c\in\mathbb{C}$. In Dirac notation, the two scalar products of $\langle\psi|$ with $\hat{A}\,|\varphi\rangle$ and $\langle\psi|\,\hat{A}$ with $|\varphi\rangle$ are by definition the same and one can write $\langle\psi|\hat{A}|\varphi\rangle$. This object is called the *matrix element* of the operator $\hat{A}$ with respect to ψ and φ. One can think of the operator $\hat{A}$ as either acting on the ket $|\varphi\rangle$ to the right or acting on the bra $\langle\psi|$ to the left. However, only the way an operator acts on a ket has been defined so far. Thus, the *adjoint* operator $\hat{A}^\dagger$ has to be defined using the one-to-one correspondence between a ket $|\psi\rangle$ and a bra $\langle\psi|$ and therefore the correspondence $|\psi'\rangle=\hat{A}\,|\psi\rangle\leftrightarrow\langle\psi'|=\langle\psi|\,\hat{A}^\dagger$.

The adjoint operator is linear because

$$\begin{aligned}
\langle\psi|=a_1\langle\psi_1|+a_2\langle\psi_2|\leftrightarrow|\psi\rangle=a_1^*|\psi_1\rangle+a_2^*|\psi_2\rangle\\
\hat{A}\,|\psi\rangle=\hat{A}\big(a_1^*\,|\psi_1\rangle+a_2^*\,|\psi_2\rangle\big)=a_1^*\,\hat{A}\,|\psi_1\rangle+a_2^*\,\hat{A}\,|\psi_2\rangle\\
a_1^*\,|\psi_1'\rangle+a_2^*\,|\psi_2'\rangle\leftrightarrow a_1\langle\psi_1'|+a_2\langle\psi_2'|=a_1\langle\psi_1|\,\hat{A}^\dagger+a_2\langle\psi_2|\,\hat{A}^\dagger\\
|\psi'\rangle\leftrightarrow\langle\psi'|=\langle\psi|\,\hat{A}^\dagger\\
\langle\psi|\,\hat{A}^\dagger=\big(a_1\langle\psi_1|+a_2\langle\psi_2|\big)\hat{A}^\dagger=a_1\langle\psi_1|\,\hat{A}^\dagger+a_2\langle\psi_2|\,\hat{A}^\dagger
\end{aligned}$$

using $\hat{A}\,|\psi_1\rangle=|\psi_1'\rangle$, $\hat{A}\,|\psi_2\rangle=|\psi_2'\rangle$ and $\hat{A}\,|\psi\rangle=|\psi'\rangle$. Other properties of the adjoint operators are

$$\langle\psi|\hat{A}^\dagger|\varphi\rangle=\langle\varphi|\hat{A}|\psi\rangle^*\qquad\left(\hat{A}^\dagger\right)^\dagger=\hat{A}\qquad\left(a\hat{A}\right)^\dagger=a^*\hat{A}^\dagger\qquad\left(\hat{A}+\hat{B}\right)^\dagger=\hat{A}^\dagger+\hat{B}^\dagger\qquad\left(\hat{A}\hat{B}\right)^\dagger=\hat{B}^\dagger\hat{A}^\dagger$$

and can be easily proven given the definition of the operator. From $\langle\psi'|\varphi\rangle=\langle\varphi|\psi'\rangle^*$ and $\hat{A}\,|\psi\rangle=|\psi'\rangle$ as well as $\langle\psi|\,\hat{A}^\dagger=\langle\psi'|$ follows $\langle\psi|\hat{A}^\dagger|\varphi\rangle=\langle\varphi|\hat{A}|\psi\rangle^*$ and from that follow further the other statements

$$\begin{aligned}
&\langle\psi|\left(\hat{A}^\dagger\right)^\dagger|\varphi\rangle=\langle\varphi|\hat{A}^\dagger|\psi\rangle^*=\left(\langle\psi|\hat{A}|\varphi\rangle^*\right)^*=\langle\psi|\hat{A}|\varphi\rangle && \Rightarrow\left(\hat{A}^\dagger\right)^\dagger=\hat{A}\\
&\langle\psi|\big(a\hat{A}\big)^\dagger|\varphi\rangle=\langle\varphi|a\hat{A}|\psi\rangle^*=a^*\langle\varphi|\hat{A}|\psi\rangle^*=a^*\langle\psi|\hat{A}^\dagger|\varphi\rangle=\langle\psi|a^*\hat{A}^\dagger|\varphi\rangle && \Rightarrow\left(a\hat{A}\right)^\dagger=a^*\hat{A}^\dagger\\
&\langle\psi|\big(\hat{A}+\hat{B}\big)^\dagger|\varphi\rangle=\langle\varphi|\big(\hat{A}+\hat{B}\big)|\psi\rangle^*=\langle\varphi|\hat{A}|\psi\rangle^*+\langle\varphi|\hat{B}|\psi\rangle^*\\
&=\langle\psi|\hat{A}^\dagger|\varphi\rangle+\langle\psi|\hat{B}^\dagger|\varphi\rangle=\langle\psi|\big(\hat{A}^\dagger+\hat{B}^\dagger\big)|\varphi\rangle && \Rightarrow\left(\hat{A}+\hat{B}\right)^\dagger=\hat{A}^\dagger+\hat{B}^\dagger\\
&|\psi'\rangle=\hat{A}\hat{B}\,|\psi\rangle=\hat{A}\,|\chi\rangle\,,\langle\psi'|=\langle\psi|\left(\hat{A}\hat{B}\right)^\dagger=\langle\chi|\,\hat{A}^\dagger=\big(\langle\psi|\,\hat{B}^\dagger\big)\hat{A}^\dagger=\langle\psi|\,\hat{B}^\dagger\hat{A}^\dagger && \Rightarrow\left(\hat{A}\hat{B}\right)^\dagger=\hat{B}^\dagger\hat{A}^\dagger
\end{aligned}$$

by using the definition of the operators, the definition of the adjoint operator and the dual space.

Given $\langle\psi|\in\mathcal{V}^*$ and $|\varphi\rangle\in\mathcal{V}$ the inner product (or scalar product) is defined as $\langle\psi|\varphi\rangle\in\mathbb{C}$ as introduced above. The question is whether $|\varphi\rangle\langle\psi|$ has a meaning, and the answer is yes. This is called the *outer product* and represents an operator because it gives a ket when applied to a ket

$$\big(|\varphi\rangle\langle\psi|\big)|\chi\rangle=|\varphi\rangle\big(\langle\psi|\chi\rangle\big)=a\,|\varphi\rangle\qquad\qquad a=\langle\psi|\chi\rangle\in\mathbb{C}$$

and this is exactly what defines an operator. One can easily check that $(|\varphi\rangle\langle\psi|)^\dagger=|\psi\rangle\langle\varphi|$.

Two subsets of operators are of special importance in quantum physics. The *Hermitian* operators have the property $\hat{A}=\hat{A}^\dagger$, and the *unitary* operators are defined by $\hat{A}^{-1}=\hat{A}^\dagger$ or by $\hat{A}^\dagger\hat{A}=\mathbb{I}$ where the operator $\mathbb{I}$ is the identity operator defined by $\mathbb{I}\,|\psi\rangle=|\psi\rangle$ for all $|\psi\rangle$.

1.3 Representations

To work with a vector space one has to represent its vectors in a convenient basis. The physics of the problem is independent of the chosen representation but the mathematics can be made much simpler by choosing a good representation. Thus, representations allow to go from the abstract space state to the concrete mathematics to solve actual quantum mechanical problems. One defines a *representation* in a state space through an *orthonormal basis* that spans the full space. A set of kets $\{|u_j\rangle\}$ is orthonormal if $\langle u_j|u_k\rangle = \delta_{jk}$ and forms a basis if there is a unique expansion

$$|\psi\rangle = \sum_j c_j \, |u_j\rangle$$

for every ket $|\psi\rangle$ in the state space where the numbers $c_j \in \mathbb{C}$ are called *expansion coefficients*.

The projection

$$\langle u_k|\psi\rangle = \langle u_k| \left(\sum_j c_j \, |u_j\rangle \right) = \sum_j c_j \, \langle u_j|u_k\rangle = c_k$$

of $|\psi\rangle$ on the basis vectors gives the expansion coefficients for an orthonormal basis. The coefficients $\{c_j\}$ are called the *representation* of $|\psi\rangle$ in the $\{|u_j\rangle\}$ basis. In the two-dimensional Euclidean space with the orthonormal basis $\{\vec{x}, \vec{y}\}$ such that $\vec{x} \cdot \vec{x} = \vec{y} \cdot \vec{y} = 1$ and $\vec{x} \cdot \vec{y} = 0$ every vector $\vec{r}$ can be represented in a similar way as $\vec{r} = a\,\vec{x} + b\,\vec{y}$ where the coefficients a and b

$$\vec{x} \cdot \vec{r} = a\,\vec{x} \cdot \vec{x} + b\,\vec{x} \cdot \vec{y} = a \qquad\qquad \vec{y} \cdot \vec{r} = a\,\vec{y} \cdot \vec{x} + b\,\vec{y} \cdot \vec{y} = b$$

are determined by projecting $\vec{r}$ onto the basis vectors.

The closure relation following from

$$|\psi\rangle = \sum_j \langle u_j|\psi\rangle \, |u_j\rangle = \sum_j |u_j\rangle \, \langle u_j|\psi\rangle = \left(\sum_j |u_j\rangle\langle u_j| \right) |\psi\rangle$$

states that

$$\sum_j |u_j\rangle\langle u_j| = \mathbb{I} \tag{1.2}$$

because the operator $\sum_j |u_j\rangle\langle u_j|$ leaves $|\psi\rangle$ invariant. The result (1.2) is also called the *resolution of the identity*.

Knowing the representation of kets, the two questions remain how bras and operators are represented. Using (1.2) and $\langle\psi| \in \mathcal{V}^*$

$$\langle\psi| = \langle\psi|\,\mathbb{I} = \langle\psi| \left(\sum_j |u_j\rangle\langle u_j| \right) = \sum_j \langle\psi|u_j\rangle \, \langle u_j| = \sum_j \langle u_j|\psi\rangle^* \, \langle u_j| = \sum_j c_j^* \, \langle u_j|$$

using the conjugation property of the scalar product. To gain the representation for operators

$$\hat{A}\,|\psi\rangle = |\psi'\rangle \qquad |\psi\rangle = \sum_j c_j \, |u_j\rangle \qquad c_j = \langle u_j|\psi\rangle \qquad |\psi'\rangle = \sum_j c'_j \, |u_j\rangle \qquad c'_j = \langle u_j|\psi'\rangle$$

is used in

$$c'_j = \langle u_j|\psi'\rangle = \langle u_j|\hat{A}|\psi\rangle = \langle u_j|\hat{A}\,\mathbb{I}|\psi\rangle = \langle u_j|\hat{A} \left(\sum_k |u_k\rangle\langle u_k| \right) |\psi\rangle = \sum_k \langle u_j|\hat{A}|u_k\rangle \, \langle u_k|\psi\rangle$$

$$|\psi'\rangle = \sum_j \left(\sum_k \langle u_j|\hat{A}|u_k\rangle \, \langle u_k|\psi\rangle \right) |u_j\rangle = \left(\sum_{jk} |u_j\rangle \, \langle u_j|\hat{A}|u_k\rangle \, \langle u_k| \right) |\psi\rangle$$

showing that

$$\hat{A} = \sum_{jk} A_{jk} \, |u_j\rangle\langle u_k| \qquad\qquad A_{jk} = \langle u_j|\hat{A}|u_k\rangle$$

meaning that the numbers A_{jk} represent the operator $\hat{A}$ in the given basis. As an example the identity operator $\mathbb{I}$ is used

$$(\mathbb{I})_{jk} = \langle u_j|\mathbb{I}|u_k\rangle = \langle u_j|u_k\rangle = \delta_{jk} \qquad \Rightarrow \qquad \mathbb{I} = \sum_{jk} (\mathbb{I})_{jk} \, |u_j\rangle\langle u_k| = \sum_{jk} \delta_{jk} \, |u_j\rangle\langle u_k| = \sum_{j} |u_j\rangle\langle u_j|$$

and this is the same as (1.2). To summarize, kets, bras and operators are represented as

$$\begin{aligned} |\psi\rangle &= \sum_j c_j \, |u_j\rangle & c_j &= \langle u_j|\psi\rangle \\ \langle\psi| &= \sum_j c_j^* \, \langle u_j| & c_j^* &= \langle\psi|u_j\rangle \\ \hat{A} &= \sum_{jk} A_{jk} \, |u_j\rangle\langle u_k| & A_{jk} &= \langle u_j|\hat{A}|u_k\rangle \end{aligned} \tag{1.3}$$

in the orthonormal basis $\{|u_j\rangle\}$.

The case of a basis $\{|u_j\rangle\}$ with a discrete index j can be generalized to a basis $\{|v_\alpha\rangle\}$ with a continuous index α. Sums become integrals and the Kronecker delta becomes a Dirac delta function. This means

$$\langle u_j|u_k\rangle = \delta_{jk} \qquad\qquad \Rightarrow \qquad\qquad \langle v_\alpha|v_\beta\rangle = \delta(\alpha - \beta)$$

$$|\psi\rangle = \sum_j c_j \, |u_j\rangle \qquad c_k = \langle u_k|\psi\rangle \qquad \Rightarrow \qquad |\psi\rangle = \int d\alpha \, c(\alpha) \, |v_\alpha\rangle \qquad c(\beta) = \langle v_\beta|\psi\rangle$$

$$\hat{A} = \sum_{jk} A_{jk} \, |u_j\rangle\langle u_k| \quad A_{jk} = \langle u_j|\hat{A}|u_k\rangle \quad \Rightarrow \quad \hat{A} = \int d\alpha \int d\beta \, A(\alpha,\beta) \, |v_\alpha\rangle\langle v_\beta| \quad A(\alpha,\beta) = \langle v_\alpha|\hat{A}|v_\beta\rangle$$

where

$$\langle v_\beta|\psi\rangle = \int d\alpha \, c(\alpha) \, \langle v_\beta|v_\alpha\rangle = \int d\alpha \, c(\alpha)\delta(v_\beta - v_\alpha) = c(\beta)$$

guarantees that delta functions are only used under an integral. The spin of a particle is an example of a discrete state space and the position is an example of a continuous state space.

1.4 Matrix Formulation of Quantum Mechanics

Matrix mechanics was the first consistent formulation of quantum theory, and it was first proposed by Heisenberg, Born and Jordan in 1925. It is equivalent to the later wave mechanics of Schrödinger, and they are unified in the state space formalism. The matrix formulation is very useful in the case of finite discrete bases.

In this formalism kets are written as column vectors, bras are written as row vectors, and operators are written as matrices

$$|\psi\rangle = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_j \\ \vdots \end{pmatrix} \qquad \langle\psi| = \begin{pmatrix} c_1^* & c_2^* & \dots & c_j^* & \dots \end{pmatrix} \qquad \hat{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1k} & \dots \\ A_{21} & A_{22} & \dots & A_{2k} & \dots \\ \vdots & \vdots & & \vdots & \\ A_{j1} & A_{j2} & \dots & A_{jk} & \dots \\ \vdots & \vdots & & \vdots & \ddots \end{pmatrix}$$

where $c_j = \langle u_j|\psi\rangle$, $c_j^* = \langle\psi|u_j\rangle$ and $A_{jk} = \langle u_j|\hat{A}|u_k\rangle$. In this formalism all manipulations of objects in quantum mechanics such as

$$\langle\psi|\varphi\rangle \qquad |\psi'\rangle = \hat{A}\,|\psi\rangle \qquad \hat{A}^\dagger \qquad |\psi\rangle\langle\varphi|$$

can be performed as

$$\langle\psi|\varphi\rangle = \left(\sum_j c_j^*\,|u_j\rangle\right)\left(\sum_k d_k\,|u_k\rangle\right) = \sum_{jk} c_j^* d_k\,\langle u_j|u_k\rangle = \sum_j c_j^* d_j = \begin{pmatrix} c_1^* & c_2^* & \dots\end{pmatrix}\begin{pmatrix} d_1\\ d_2\\ \vdots\end{pmatrix}$$

$$c_j' = \langle u_j|\psi'\rangle = \langle u_j|\hat{A}|\psi\rangle = \langle u_j|\hat{A}\left(\sum_k |u_k\rangle\langle u_k|\right)|\psi\rangle = \sum_k \langle u_j|\hat{A}|u_k\rangle\,\langle u_k|\psi\rangle = \sum_k A_{jk}c_k$$

$$= \begin{pmatrix} A_{11} & A_{12} & \dots\\ A_{21} & A_{22} & \dots\\ \vdots & \vdots & \end{pmatrix}\begin{pmatrix} c_1\\ c_2\\ \vdots\end{pmatrix} = \begin{pmatrix} c_1'\\ c_2'\\ \vdots\end{pmatrix}$$

$$(\hat{A}^\dagger)_{jk} = \langle u_j|\hat{A}^\dagger|u_k\rangle = \langle u_k|\hat{A}|u_j\rangle^* = A_{kj}^*$$

$$|\psi\rangle\langle\varphi| = \left(\sum_j c_j\,|u_j\rangle\right)\left(\sum_k d_k^*\,\langle u_k|\right) = \sum_{jk} c_j d_k^*\,|u_j\rangle\langle u_k| = \begin{pmatrix} c_1\\ c_2\\ \vdots\end{pmatrix}\begin{pmatrix} d_1 & d_2 & \dots\end{pmatrix} = \begin{pmatrix} c_1 d_1^* & c_1 d_2^* & \dots\\ c_2 d_1^* & c_2 d_2^* & \dots\\ \vdots & \vdots & \end{pmatrix}$$

according to the rules of matrix multiplication, and this shows that the adjoint matrix is the transpose conjugate matrix.

Changing from one basis $\{|u_j\rangle\}$ to another basis $\{|v_k\rangle\}$ is an essential process in matrix mechanics. With $c_j = \langle u_j|\psi\rangle$ and $d_k = \langle v_k|\psi\rangle$ as the coefficients in the two bases one gets

$$d_k = \langle v_k|\psi\rangle = \langle v_k|\mathbb{I}|\psi\rangle = \langle v_k|\left(\sum_j |u_j\rangle\langle u_j|\right)|\psi\rangle = \sum_j \langle v_k|u_j\rangle\,\langle u_j|\psi\rangle = \sum_j S_{kj}c_j$$

where S_{kj} is defined as $\langle v_k|u_j\rangle$ and is called the *overlap matrix*. Thus, changing from $\{|u_j\rangle\}$ to $\{|v_k\rangle\}$ uses $S_{kj} = \langle v_k|u_j\rangle$. Changing from $\{|v_k\rangle\}$ to $\{|u_j\rangle\}$ can be calculated similarly and shows that the overlap matrix is S_{kj}^* because $\langle u_j|v_k\rangle = \langle v_k|u_j\rangle^* = S_{kj}^*$.

Operators change when going from $\{|u_j\rangle\}$ to $\{|v_k\rangle\}$ as

$$A_{jm}^u = \langle u_j|\hat{A}|u_m\rangle \qquad\qquad A_{kn}^v = \langle v_k|\hat{A}|v_n\rangle$$

and transform as

$$A_{kn}^v = \sum_{jm} \langle v_k|u_j\rangle\,\langle u_j|\hat{A}|u_m\rangle\,\langle u_m|v_n\rangle = \sum_{jm} S_{kj}A_{jm}^u S_{nm}^*$$

as can be shown using the resolution of the identity (1.2) twice.

1.5 Eigenvalues and Eigenvectors of Operators

Operators are the mathematical objects that allow to represent physical observables in quantum mechanics. If one measures a physical observable the only possible outcome of this measurement is one of the eigenvalues of the associated operator. After one has measured this property, the state of the system is the eigenstate of the operator corresponding to the measured eigenvalue.

Postulate III: The result of a measurement of a physical quantity is one of the *eigenvalues* of the associated observable.

An operator acts as $\hat{A}\,|\psi\rangle = |\psi'\rangle$ and therefore takes a ket and returns a ket. If the only change for a specific ket is that the ket is scaled by some factor such that $\hat{A}\,|\psi\rangle = \lambda\,|\psi\rangle$ with $\lambda \in \mathbb{C}$. The equation

$$\hat{A}\,|\psi\rangle = \lambda\,|\psi\rangle \qquad\qquad \lambda \in \mathbb{C} \tag{1.4}$$

is called the *eigenvalue equation* for the operator $\hat{A}$. A ket $|\psi\rangle$ satisfying equation (1.4) is called a *eigenstate*, *eigenket* or *eigenvector* of operator $\hat{A}$, and λ is called an *eigenvalue* of $\hat{A}$. The eigenvectors are those special directions in a vector space which the operator does not change. The set of eigenvalues $\{\lambda\}$ is the *spectrum* of $\hat{A}$.

The eigenvalue equation (1.4) is often written as $\hat{A}\,|\lambda\rangle = \lambda\,|\lambda\rangle$ in quantum mechanics. In this convenient notation the eigenstate $|\lambda\rangle$ is labeled by the corresponding eigenvalue λ.

One property of eigenstates is that they are only defined up to a constant because $|\psi'\rangle = \alpha\,|\psi\rangle$ for $\alpha \in \mathbb{C}$ is also an eigenstate of $\hat{A}$ if $|\psi\rangle$ is an eigenstate of $\hat{A}$ as

$$\hat{A}\,|\psi'\rangle = \hat{A}(\alpha\,|\psi\rangle) = \alpha(\hat{A}\,|\psi\rangle) = \alpha(\lambda\,|\psi\rangle) = \lambda(\alpha\,|\psi\rangle) = \lambda\,|\psi'\rangle$$

shows. This means that the eigenstate is a direction in the state space and any state in that direction independent of its length is an eigenstate of this operator with the same eigenvalue. Thus, one usually uses *normalized* eigenstates by choosing α such that $\langle\psi|\psi\rangle = 1$. Because the state space is a complex vector space eigenstates are only defined up to a *global phase*. Given $\langle\psi|\psi\rangle = 1$ for an eigenstate $|\psi\rangle$ and $|\psi'\rangle = e^{i\theta}\,|\psi\rangle$ where $e^{i\theta} = \alpha$ is just a complex number and $|\psi'\rangle$ is therefore also an eigenstate. It is also normalized because $\langle\psi'|\psi'\rangle = \langle\psi|e^{-i\theta}e^{i\theta}|\psi\rangle = 1$. Thus, there is an extra degree of freedom coming from this global phase, and whatever quantum mechanics tells about nature is independent of multiplying a ket by a global phase. This ambiguity is actually true for any state and not only for eigenstates.

Another important property of eigenstate is related to degeneracy of eigenvalues. An eigenvalue is called *non-degenerate* if it is the only eigenstate for this eigenvalue (up to a constant), and an eigenvalue is called *degenerate* if there are more than one linearly independent eigenstates for this eigenvalue. If the eigenvalue λ is n-fold degenerate then one writes $\hat{A}\,|\psi^j\rangle = \lambda\,|\psi^j\rangle$ for $j = 1, 2, ..., n$ where $|\psi^j\rangle$ are linearly independent. The state

$$|\psi\rangle = \sum_j c_j\,|\psi^j\rangle \qquad\qquad c_j \in \mathbb{C}$$

is also an eigenstate for the eigenvalue λ as

$$\hat{A}\,|\psi\rangle = \hat{A}\left(\sum_j c_j\,|\psi^j\rangle\right) = \sum_j c_j\,\hat{A}\,|\psi^j\rangle = \sum_j c_j\,\lambda\,|\psi^j\rangle = \lambda\left(\sum_j c_j\,|\psi^j\rangle\right) = \lambda\,|\psi\rangle$$

proves. This shows that any linear combination of eigenstates in an n-fold degenerate subspace is also an eigenstate with the same eigenvalue.

The remaining question is how to find the eigenstates and eigenvalues of an operator $\hat{A}$. One can take the identity operator $\mathbb{I}$ with $\mathbb{I}\,|\psi\rangle = |\psi\rangle$ for all $|\psi\rangle$ as a simple case to see that any state $|\psi\rangle$ is an eigenstate of $\mathbb{I}$ with $\lambda = 1$ such that this eigenvalue is infinitely degenerate. One can determine the eigenstates and eigenvalues in this case by just looking at how the operator acts, and the same is possible for projection operators and the parity operator.

This is, however, not possible for other operators. A general approach starts from a basis $\{|u_j\rangle\}$ with the property to be *orthonormal* such that $\langle u_j|u_k\rangle = \delta_{jk}$. With

$$|\psi\rangle = \sum_j c_j\,|u_j\rangle \qquad\qquad \hat{A} = \sum_{jk} A_{jk}\,|u_j\rangle\langle u_k|$$

one gets

$$\lambda\,\langle u_j|\psi\rangle = \langle u_j|\hat{A}|\psi\rangle = \langle u_j|\hat{A}\,\mathbb{I}|\psi\rangle = \langle u_j|\hat{A}\left(\sum_k |u_k\rangle\langle u_k|\right)|\psi\rangle = \sum_k \langle u_j|\hat{A}|u_k\rangle\,\langle u_k|\psi\rangle$$

and finally

$$\sum_k (A_{jk} - \lambda\,\delta_{jk})\,c_k = 0 \qquad\qquad (\underline{\underline{A}} - \lambda\,\underline{\underline{\mathbb{I}}})\underline{c} = \underline{0} \tag{1.5}$$

because $\langle u_j|\psi\rangle = c_j$. This equation is equivalent to the original eigenvalue equation above but represented in the basis $\{|u_j\rangle\}$. Eigenvalues are the same in any representation because (1.4) is independent of the representation. The task of finding the eigenstates and eigenvalues of the operator $\hat{A}$ means finding λ and c_k, and (1.5) shown in two forms is a linear homogeneous system of equations for λ and c_k. The task is called matrix diagonalization. In an N-dimensional state space there are N equations of N unknowns.

Solving the characteristic equation $\det(\underline{\underline{A}} - \lambda\,\underline{\underline{\mathbb{I}}}) = 0$ allows to determine λ. It is an N-order polynomial equation for λ and has N solutions called roots which can be real or imaginary numbers. They can be unique solutions called simple roots or repeated solutions called multiple roots.

Solving the equation $(\underline{\underline{A}} - \lambda\,\underline{\underline{\mathbb{I}}})\underline{c} = \underline{0}$ gives one eigenvector for simple roots but for multiple roots the situation is more difficult. If a multiple root with multiplicity p has p linearly independent eigenvectors then one can diagonalize the matrix, but if there are fewer than p linearly independent eigenvectors then the matrix cannot be diagonalized. Because one is only interested in Hermitian and unitary operators in quantum physics this problem cannot occur.

As an example the operator $\hat{A}$ and state $|\psi\rangle$ in matrix form

$$\hat{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad\qquad |\psi\rangle = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

are used and gives

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \lambda \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

as the eigenvalue equation. The characteristic equation leads to

$$\det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = 0 \qquad \Rightarrow \qquad (-\lambda)^2 - (-1) = 0 \qquad \Rightarrow \qquad \lambda^2 = -1 \qquad \Rightarrow \qquad \lambda = \pm i$$

with two solutions $\lambda = +i$ and $\lambda = -i$. The eigenvector for $\lambda = +i$ is

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = i \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \qquad \Rightarrow \qquad \begin{pmatrix} -c_2 \\ c_1 \end{pmatrix} = i \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \qquad \Rightarrow \qquad \left| \begin{array}{c} -c_2 = ic_1 \\ c_1 = ic_2 \end{array} \right|$$

where both equations are equivalent because $i(c_1 = ic_2) \Rightarrow ic_1 = -c_2$. A similar calculation for $\lambda = -i$ gives the two normalized eigenvectors

$$\lambda = +i \Rightarrow |\psi\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix} \qquad\qquad \lambda = -i \Rightarrow |\psi\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix}$$

for these two eigenvalues. (Note that these eigenvectors can be multiplied by a phase and still represent the same state.)

1.6 Hermitian Operators

The subset of operators consisting of the Hermitian operators allows to describe physical observables. A *Hermitian* operator $\hat{A}$ is an operator that is equal to its adjoint such that $\hat{A} = \hat{A}^\dagger$. Because of the duality $\hat{A}\,|\psi\rangle = |\psi'\rangle \Leftrightarrow \langle\psi|\,\hat{A}^\dagger = \langle\psi'|$ this means for Hermitian operators $\hat{A}\,|\psi\rangle = |\psi'\rangle$ and $\langle\psi|\,\hat{A} = \langle\psi'|$. Hermitian operators have two important properties.

The first property is that the eigenvalues of a Hermitian operator are real numbers. This is important since the Hermitian operators represent physical observables in quantum mechanics and measurements

result in real numbers. Starting from the eigenvalue equation (1.4) for a Hermitian operator $\hat{A}\left|\psi\right\rangle = \lambda\left|\psi\right\rangle$ one gets $\left\langle\psi\right|\hat{A}\left|\psi\right\rangle = \lambda\left\langle\psi|\psi\right\rangle$. Starting on the other hand from the dual equation $\left\langle\psi\right|\hat{A}^{\dagger} = \lambda^{*}\left\langle\psi\right|$ which is $\left\langle\psi\right|\hat{A} = \lambda^{*}\left\langle\psi\right|$ for a Hermitian operator one gets $\left\langle\psi\right|\hat{A}\left|\psi\right\rangle = \lambda^{*}\left\langle\psi|\psi\right\rangle$. This leads to $\lambda\left\langle\psi|\psi\right\rangle = \lambda^{*}\left\langle\psi|\psi\right\rangle$ because the left sides of both resulting equations are the same. Positivity means that $\left\langle\psi|\psi\right\rangle \geq 0$ and $\left\langle\psi|\psi\right\rangle = 0$ only for $\left|\psi\right\rangle = 0$ and it follows $\lambda = \lambda^{*}$ or, in other words, λ is a real number $\lambda \in \mathbb{R}$.

The second property is that the eigenstates of a Hermitian operator can be chosen to form an orthonormal basis. One can therefore always write an arbitrary quantum state in the representation given by the eigenstates of the physical observable. Let $\left|\psi\right\rangle$ and $\left|\varphi\right\rangle$ be two eigenstates of the Hermitian operator $\hat{A}$ with $\hat{A}\left|\psi\right\rangle = \lambda\left|\psi\right\rangle$ and $\hat{A}\left|\varphi\right\rangle = \mu\left|\varphi\right\rangle$.

First it is assumed that $\lambda \neq \mu$. From $\hat{A}\left|\psi\right\rangle = \lambda\left|\psi\right\rangle$ follows $\left\langle\varphi\right|\hat{A}\left|\psi\right\rangle = \lambda\left\langle\varphi|\psi\right\rangle$ and from $\left\langle\varphi\right|\hat{A} = \mu\left\langle\varphi\right|$ (because $\hat{A} = \hat{A}^{\dagger}$ and $\mu \in \mathbb{R}$) follows $\left\langle\varphi\right|\hat{A}\left|\psi\right\rangle = \mu\left\langle\varphi|\psi\right\rangle$. Because the two left sides are equal it follows that $\lambda\left\langle\varphi|\psi\right\rangle = \mu\left\langle\varphi|\psi\right\rangle$ or $(\lambda - \mu)\left\langle\varphi|\psi\right\rangle = 0$. Thus, $\left\langle\varphi|\psi\right\rangle = 0$ because $\lambda - \mu \neq 0$, and this means that two eigenstates with different eigenvalues are orthogonal. More general, this means that all eigenstates are mutually orthogonal if all eigenvalues are different, and one can normalize them in order to build an orthonormal basis.

If, however, $\lambda = \mu$ then the eigenvalues are degenerate. For n degenerate eigenvalues the eigenvalue equation becomes $\hat{A}\left|\psi^{j}\right\rangle = \lambda\left|\psi^{j}\right\rangle$ for $j = 1, ..., n$. Because any linear combination of these n eigenstates $\left|\psi^{j}\right\rangle$ is also an eigenstate of λ the set of eigenstates of λ is a subspace of the state space, and because for an n-fold degenerate eigenvalue there are n linearly independent eigenstates this subspace must therefore have dimension n.

The eigenstates for different eigenvalues are orthogonal but one only knows that eigenstates belonging to degenerate eigenvalues can be chosen linearly independent. The Gram-Schmidt orthonormalization algorithm shows that one can always find an orthonormal basis for the subspace of eigenstates for an eigenvalue. The initial linearly independent but not necessarily orthogonal set of eigenstates is $\{\left|\psi^{j}\right\rangle\}$ for $j = 1, ..., n$, and the goal is to find $\{\left|\varphi^{j}\right\rangle\}$ for $j = 1, ..., n$ with $\left\langle\varphi^{j}|\varphi^{k}\right\rangle = \delta_{jk}$. The algorithm works as

$$\left|\varphi^{1}\right\rangle = \frac{\left|\psi^{1}\right\rangle}{\sqrt{\left\langle\psi^{1}|\psi^{1}\right\rangle}}$$

$$\left|\chi^{2}\right\rangle = \left|\psi^{2}\right\rangle + \alpha\left|\varphi^{1}\right\rangle = \left|\psi^{2}\right\rangle - \left|\varphi^{1}\right\rangle\left\langle\varphi^{1}|\psi^{2}\right\rangle \Rightarrow \left|\varphi^{2}\right\rangle = \frac{\left|\chi^{2}\right\rangle}{\sqrt{\left\langle\chi^{2}|\chi^{2}\right\rangle}}$$

for the first two eigenstates using that from $\left\langle\varphi^{1}|\psi^{2}\right\rangle + \alpha\left\langle\varphi^{1}|\varphi^{1}\right\rangle = 0$ follows $\alpha = -\left\langle\varphi^{1}|\psi^{2}\right\rangle$. The general idea is to subtract from $\left|\psi^{j}\right\rangle$ a linear combination of $\left|\varphi^{1}\right\rangle, ..., \left|\varphi^{j-1}\right\rangle$ to make it orthogonal to all the previously determined $\left|\varphi^{k}\right\rangle$.

Hermitian operators in the matrix formulation of quantum mechanics are Hermitian matrices. Because the adjoint matrix is the transpose conjugate matrix one gets

$$\hat{A} \Rightarrow A_{jk} \qquad\qquad \hat{A}^{\dagger} \Rightarrow A_{kj}^{*}$$

for any operator $\hat{A}$. This means $A_{jk} = A_{kj}^{*}$ for a Hermitian matrix. The diagonal elements of a Hermitian matrix are real numbers because $j = k$ means $A_{jj} = A_{jj}^{*}$. Thus, the general Hermitian matrix can be written as

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & \dots \\ A_{21} & A_{22} & A_{23} & \dots \\ A_{31} & A_{32} & A_{33} & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix} = \begin{pmatrix} A_{11} \in \mathbb{R} & A_{12} \in \mathbb{C} & A_{13} \in \mathbb{C} & \dots \\ A_{12}^{*} \in \mathbb{C} & A_{22} \in \mathbb{R} & A_{23} \in \mathbb{C} & \dots \\ A_{13}^{*} \in \mathbb{C} & A_{23}^{*} \in \mathbb{C} & A_{33} \in \mathbb{R} & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

and the real elements in the diagonal and and the complex elements above the diagonal completely determine the matrix.

The Hermitian matrix for $\hat{A}$ in the basis of eigenstates with $\hat{A}\left|\varphi_{j}\right\rangle = \lambda_{j}\left|\varphi_{j}\right\rangle$ and $\left\langle\varphi_{j}|\varphi_{k}\right\rangle = \delta_{jk}$ is diagonal because $A_{jk} = \left\langle\varphi_{j}\right|\hat{A}\left|\varphi_{k}\right\rangle = \lambda_{k}\left\langle\varphi_{j}|\varphi_{k}\right\rangle = \lambda_{k}\delta_{jk}$. The diagonal matrix elements are the eigenvalues of $\hat{A}$ and the other matrix elements vanish. To summarize, the matrix associated with a Hermitian operator $\hat{A}$ is diagonal when written in the basis spanned by the eigenstates of $\hat{A}$, and the diagonal elements are the eigenvalues of $\hat{A}$.

1.7 Compatible Observables

Compatible observables are observables that commute. The significance for physics is that compatible observables can be measured simultaneously. If two observables such as position and momentum in quantum mechanics do not commute then they cannot be measured simultaneously.

The fact that two observables $\hat{A}$ and $\hat{B}$ are compatible can be expressed as $[\hat{A},\hat{B}]=0$ using (1.1). The important property of compatible observables is that if $[\hat{A},\hat{B}]=0$ then it is possible to find a common set of eigenstates of $\hat{A}$ and $\hat{B}$ that form a basis in state space. The observable $\hat{A}$ can be diagonalized such that the first g_1 eigenvalues in the diagonal are λ_1, the next g_2 eigenvalues are λ_2 and so on. Thus, eigenvalues can be degenerate. Because the g_j states in the basis corresponding to λ_j can be selected arbitrarily within the subspace with dimension g_j as long as they are orthonormal, they may not be eigenstates of $\hat{B}$. The matrix for $\hat{B}$ is only block-diagonal and this means that there is a block with $g_1 \times g_1$ elements not being zero, and the same is true for any other degenerate eigenvalue. However, it is possible to choose a basis within the subspace of dimension g_j belonging to λ_j such that also the matrix for $\hat{B}$ is diagonal. This is now proven in detail.

In a first step it is shown that an eigenvalue of $\hat{A}$ is also an eigenvalue of $\hat{B}$. Given $\hat{A}\,|\psi\rangle=\lambda\,|\psi\rangle$ as the eigenvalue equation (1.4) and $|\chi\rangle=\hat{B}\,|\psi\rangle$ then

$$\hat{A}\,|\chi\rangle=\hat{A}\hat{B}\,|\psi\rangle=\hat{B}\hat{A}\,|\psi\rangle=\lambda\big(\hat{B}\,|\psi\rangle\big)=\lambda\,|\chi\rangle$$

using $\hat{A}\hat{B}=\hat{B}\hat{A}$ shows that λ is also an eigenvalue of $\hat{B}$. If λ is non-degenerate then $|\psi\rangle$ and $|\chi\rangle$ are proportional, and the proportionality factor μ with $|\chi\rangle=\hat{B}\,|\psi\rangle=\mu\,|\psi\rangle$ gives just the eigenvalue equation for $\hat{B}$. Thus, if $[\hat{A},\hat{B}]=0$ and the eigenvalue spectra of $\hat{A}$ and $\hat{B}$ are non-degenerate then their eigenstates are the same.

The degenerate case is slightly more complicated. Starting from $\hat{A}\,|\psi_m^j\rangle$ for $j=1,...,g_m$ as the set of all g_m eigenstates associated with the eigenvalue λ_m where the sum of all the numbers g_m gives the dimension of the state space, all the eigenstates of $\hat{A}$ are assumed to be orthonormal (using the Gram-Schmidt orthonormalization algorithm) such that $\langle\psi_m^j|\psi_n^k\rangle=\delta_{mn}\delta_{jk}$. The matrix for the operator $\hat{A}$ in the basis of its eigenstates is diagonal.

With $B_{mn}^{jk}=\langle\psi_m^j|\hat{B}|\psi_n^k\rangle$ the case for $m\neq n$ gives

$$\begin{aligned}
\langle\psi_m|\,\hat{A}=\lambda_m\hat{A} \quad &\Rightarrow \quad \langle\psi_m|\hat{A}\hat{B}|\psi_n\rangle=\lambda_m\,\langle\psi_m|\hat{B}|\psi_n\rangle\\
\hat{A}\,|\psi_n\rangle=\lambda_n\,|\psi_n\rangle \quad &\Rightarrow \quad \langle\psi_m|\hat{B}\hat{A}|\psi_n\rangle=\lambda_n\,\langle\psi_m|\hat{B}|\psi_n\rangle\\
&\Rightarrow \quad \langle\psi_m|\hat{B}|\psi_n\rangle=0
\end{aligned}$$

because $\langle\psi_m|\hat{A}\hat{B}|\psi_n\rangle=\langle\psi_m|\hat{B}\hat{A}|\psi_n\rangle$ and $\lambda_m\neq\lambda_n$ follows $\langle\psi_m^j|\hat{B}|\psi_n^k\rangle=0$. This case is basically the same as the non-degenerate case.

The case for $m=n$ is different because $\lambda_m=\lambda_n$, and the matrix of $\hat{B}$ in the eigenstates of $\hat{A}$ is a the block-diagonal matrix $\underline{\underline{B}}$. In the $g_m\times g_m$ block $\underline{\underline{B}}_m$ (a submatrix of $\underline{\underline{B}}$) corresponding to λ_m, the eigenstates of $\hat{B}$ are linear combinations of the eigenstates $|\psi_m^j\rangle$ for $j=1,...,g_m$. Because $\underline{\underline{B}}$ is Hermitian also $\underline{\underline{B}}_m$ is obviously Hermitian, and it therefore can be diagonalized for $\hat{B}$. Since any linear combination of eigenstates of $\hat{A}$ for the same eigenvalue is an eigenstate of $\hat{A}$ the eigenstates of $\hat{B}$ after diagonalizing the blocks are still eigenstates of $\hat{A}$. Thus, one can always find a common basis of orthonormal eigenstates for two Hermitian operators $\hat{A}$ and $\hat{B}$ if the commutator $[\hat{A},\hat{B}]$ vanishes.

The inverse of this statement is easy to prove. If there is a basis of eigenstates common to $\hat{A}$ and $\hat{B}$, then $[\hat{A},\hat{B}]=0$. For any common eigenstate $|\psi\rangle$ the two equations

$$\hat{A}\hat{B}\,|\psi\rangle=\mu\hat{A}\,|\psi\rangle=\lambda\mu\,|\psi\rangle \qquad\qquad \hat{B}\hat{A}\,|\psi\rangle=\lambda\hat{B}\,|\psi\rangle=\lambda\mu\,|\psi\rangle$$

hold for some eigenvalues λ and μ. Because any state can be written as a linear combination of the eigenstates forming a basis, $(\hat{A}\hat{B}-\hat{B}\hat{A})\,|\varphi\rangle=0$ for all $|\varphi\rangle$ in the state space. Actually, the simple reason for this inverse statement boils down to the fact that the multiplication of numbers is commutative. Two observables can therefore be measured simultaneously if and only if they commute.

1.8 Projection Operators

Projection operators project one quantum state onto another. One case where this is useful is when one measures a property of a quantum particle. Quantum mechanics tells that the state of the particle collapses onto a different state, and this collapse can be mathematically described by projection operators.

Given a state $|\psi\rangle$ that is normalized such that $\langle\psi|\psi\rangle = 1$, the *projection* operator is defined as the outer product $\hat{P}_\psi = |\psi\rangle\langle\psi|$. The action of this operator is

$$\hat{P}_\psi |\varphi\rangle = \big(|\psi\rangle\langle\psi| \big) |\varphi\rangle = |\psi\rangle \langle\psi|\varphi\rangle = c\,|\psi\rangle \qquad\qquad c = \langle\psi|\varphi\rangle$$

on an arbitrary state $|\varphi\rangle$. The projection operator associated with the ket $|\psi\rangle$ takes a ket $|\varphi\rangle$ and returns another ket proportional to $|\psi\rangle$. The proportionality constant c is given by the overlap between the initial state $|\varphi\rangle$ and the state $|\psi\rangle$ that defines the projection operator. Thus, the projection operator projects an arbitrary state onto the reference state $|\psi\rangle$.

The first property of projection operators is that they are *idempotent* as

$$\hat{P}_\psi^2 = \big(|\psi\rangle\langle\psi| \big)\big(|\psi\rangle\langle\psi| \big) = |\psi\rangle \langle\psi|\psi\rangle \langle\psi| = |\psi\rangle\, 1\, \langle\psi| = |\psi\rangle\langle\psi| = \hat{P}_\psi$$

shows using the fact that $|\psi\rangle$ is normalized. This makes sense because projecting the projection on $|\psi\rangle$ again on $|\psi\rangle$ does not change anything. The second property of projection operators is that they are Hermitian because

$$\hat{P}_\psi = \big(|\psi\rangle\langle\psi| \big)^\dagger = |\psi\rangle\langle\psi| = \hat{P}_\psi$$

due to the fact that the adjoint of an outer product simply reverses the order.

The eigenvalues and eigenstates of the projection operator $\hat{P}_\psi$ satisfy

$$\hat{P}_\psi |\lambda\rangle = \lambda\,|\lambda\rangle \Rightarrow |\psi\rangle \langle\psi|\lambda\rangle = \lambda\,|\lambda\rangle \qquad\qquad c\,|\psi\rangle = \lambda\,|\lambda\rangle \qquad\qquad c = \langle\psi|\lambda\rangle$$

with the two solutions $|\lambda\rangle = |\psi\rangle$ and $\lambda = 1$ or $\langle\psi|\lambda\rangle = 0$ and $\lambda = 0$. In the first solution is the direction of the eigenvector parallel to $|\psi\rangle$ and in the second solution orthogonal to $|\psi\rangle$.

An important property of projection operators is that one can write any state as the sum of a state parallel to $|\psi\rangle$ and a state orthogonal to $|\psi\rangle$ because

$$|\varphi\rangle = \mathbb{I}\,|\varphi\rangle = \mathbb{I}\,|\varphi\rangle + \hat{P}_\psi\,|\varphi\rangle - \hat{P}_\psi\,|\varphi\rangle = \hat{P}_\psi\,|\varphi\rangle + (\mathbb{I} - \hat{P}_\psi)\,|\varphi\rangle$$

splits $|\varphi\rangle$ into two eigenstates of $\hat{P}_\psi$. The first part $\hat{P}_\psi\,|\varphi\rangle$ is an eigenstate with eigenvalue $\lambda = 1$ because $\hat{P}_\psi[\hat{P}_\psi\,|\varphi\rangle] = \hat{P}_\psi^2\,|\varphi\rangle = \hat{P}_\psi\,|\varphi\rangle$ and the second part $(\mathbb{I} - \hat{P}_\psi)\,|\varphi\rangle$ is an eigenstate with eigenvalue $\lambda = 0$ because $\hat{P}_\psi[(\mathbb{I} - \hat{P}_\psi)\,|\varphi\rangle] = (\hat{P}_\psi - \hat{P}_\psi^2)\,|\varphi\rangle = (\hat{P}_\psi - \hat{P}_\psi)\,|\varphi\rangle = 0$. Thus, any state $|\varphi\rangle$ can be written as

$$|\varphi\rangle = \hat{P}_\psi\,|\varphi\rangle + \big(\mathbb{I} - \hat{P}_\psi\big)\,|\varphi\rangle \tag{1.6}$$

where $\hat{P}_\psi\,|\varphi\rangle$ is parallel to $|\psi\rangle$ and $(\mathbb{I} - \hat{P}_\psi)\,|\varphi\rangle$ is orthogonal to $|\psi\rangle$.

Given are two vector spaces U_1 and U_2 where both are subspace of another vector space V. If one can always write any vector $|v\rangle \in V$ as a sum of a vector $|u_1\rangle \in U_1$ and a vector $|u_2\rangle \in U_2$ then U_1 and U_2 span V written as $V = U_1 + U_2$. If this decomposition is unique such that $U_1 \cap U_2 = \{0\}$ then is V given by the so-called direct sum $V = U_1 \oplus U_2$ and U_1 and U_2 are called complementary subspaces. Because any $|\psi\rangle \in V$ can be decomposed according to (1.6) into $\hat{P}_\psi\,|\varphi\rangle \in V_1$ and $(\mathbb{I} - \hat{P}_\psi)\,|\varphi\rangle \in V_0$ every projection operator splits the vector space V into two complementary subspaces such that $V = V_1 \oplus V_0$.

The n vectors $|u_1\rangle, ..., |u_n\rangle$ build a subspace of a vector space with the orthonormal basis $\{|u_j\rangle\}$, and the operator defined as

$$\hat{P}_n = \sum_{j=1}^{n} |u_j\rangle\langle u_j| \tag{1.7}$$

is a projection operator because

$$\hat{P}_n^2 = \left(\sum_{j=1}^{n} |u_j\rangle\langle u_j|\right)\left(\sum_{k=1}^{n} |u_k\rangle\langle u_k|\right) = \sum_{j=1}^{n}\sum_{k=1}^{n} |u_j\rangle \langle u_j|u_k\rangle \langle u_k| = \sum_{j=1}^{n}\sum_{k=1}^{n} |u_j\rangle \delta_{jk} \langle u_k| = \sum_{j=1}^{n} |u_j\rangle\langle u_j| = \hat{P}_n$$

is satisfied. Therefore, the projection of an arbitrary vector $|\varphi\rangle$ is

$$\hat{P}_n = \sum_{j=1}^{n} |u_j\rangle \langle u_j|\varphi\rangle$$

and is a vector in the subspace spanned by the basis vectors $|u_1\rangle, ..., |u_n\rangle$.

1.9 Measurement in Quantum Mechanics

The measurement process in quantum mechanics is very different from the one in classical physics due to the probabilistic nature of quantum mechanics. It is therefore impossible to predict the outcome of a quantum mechanical measurement. A physical quantity $\mathcal{A}$ corresponds to an observable $\hat{A}$ with the eigenvalue equation $\hat{A}\,|u_j\rangle = \lambda_j\,|u_j\rangle$. According to postulate III the result of a measurement of a physical quantity is one of the eigenvalues of the associated observable. This postulate only tells that one gets one of the eigenvalues of $\hat{A}$ if one measures $\hat{A}$ in state $|\psi\rangle$, but it does not tell which eigenvalue one will get.

Postulate IV: The measurement of $\mathcal{A}$ in a system in normalized state $|\psi\rangle$ gives eigenvalue λ_j with probability $P(\lambda_j) = |\langle u_j|\psi\rangle|^2$.

This means that with N copies of $|\psi\rangle$ one gets the result λ_j of a measurement p_j times, and

$$\lim_{N\to\infty} \frac{p_j}{N} \to P(\lambda_j)$$

relates p_j to $P(\lambda_j)$ for large numbers N. For $N = 1$ only the probabilities for the different λ_j are known, but for very large N on the other hand the rates for the different λ_j are known.

With $\psi = \sum_j c_j\,|u_j\rangle$ and $c_j = \langle u_j|\psi\rangle$ the probability can be written as $P(\lambda_j) = |c_j|^2$. This means that one expands $|\psi\rangle$ in the basis of eigenstates of $\hat{A}$ in order to measure property $\hat{A}$ in state $|\psi\rangle$, and the expansion coefficients c_j tell the relative contribution of λ_j. A very special case is when $|\psi\rangle$ is one of the eigenstates $|u_j\rangle$, and the result of a measurement is certain because of $c_j = 1$ and $P(\lambda_j) = |c_j|^2 = 1$.

The implicit assumptions of postulate IV

1. $\langle\psi|\psi\rangle = 1$
2. λ_j is non-degenerate
3. λ_j is discrete

can be relaxed.

From $P(\lambda_j) = |c_j|^2$, $\sum_j P(\lambda_j) = 1$ and

$$\langle\psi|\psi\rangle = \left(\sum_j c_j^*\,\langle u_j|\right)\left(\sum_k c_k\,|u_k\rangle\right) = \sum_{jk} c_j^* c_k\,\langle u_j|u_k\rangle = \sum_j |c_j|^2$$

follows the generalized statement

$$\langle\psi|\psi\rangle = C \qquad \Rightarrow \qquad P(\lambda_j) = \frac{1}{C}|\langle u_j|\psi\rangle|^2 = \frac{|\langle u_j|\psi\rangle|^2}{\langle\psi|\psi\rangle} \tag{1.8}$$

in case $|\psi\rangle$ is not normalized. Thus, the probability $P_\psi(\lambda_j)$ can be determined even if the length of $|\psi\rangle$ is $r \neq 1$, and the probability is also independent of a global phase as discussed above. If $P_\psi(\lambda_j)$ is the probability given by (1.8) then the probability of $|\psi'\rangle = r\,e^{i\theta}\,|\psi\rangle$ is

$$P_{\psi'}(\lambda_j) = \frac{|\langle u_j|\psi'\rangle|^2}{\langle\psi'|\psi'\rangle} = \frac{\left|\langle u_j|r\,e^{i\theta}|\psi\rangle\right|^2}{\langle\psi|r\,e^{-i\theta}\,r\,e^{i\theta}|\psi\rangle} = \frac{r^2\,|\langle u_j|\psi\rangle|^2}{r^2\,\langle\psi|\psi\rangle} = \frac{|\langle u_j|\psi\rangle|^2}{\langle\psi|\psi\rangle} = P_\psi(\lambda_j)$$

as expected.

The eigenvalue equation in the case of degenerate eigenvalues is

$$\hat{A}\,|u_m^j\rangle = \lambda_m\,|u_m^j\rangle \qquad\qquad j = 1, ..., g_m$$

and the g_m eigenstates for the eigenvalue λ_m form a subspace of dimension g_m where it is assumed that $\{|u_m^j\rangle\}$ is an orthonormal basis. An arbitrary state $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_m \sum_{j=1}^{g_m} c_m^j\,|u_m^j\rangle \qquad\qquad c_m^j = \langle u_m^j|\psi\rangle$$

such that the probability becomes

$$P(\lambda_m) = \sum_{j=1}^{g_m} \left|\langle u_m^j|\psi\rangle\right|^2 = \sum_{j=1}^{g_m} \left|c_m^j\right|^2$$

for the eigenvalue λ_m. Using the projection operator on a subspace defined as (1.7) and replacing $|\psi\rangle$ by

$$|\psi_m\rangle = \hat{P}_m\,|\psi\rangle = \sum_{j=1}^{g_m} |u_m^j\rangle\,\langle u_m^j|\psi\rangle = \sum_{j=1}^{g_m} c_m^j\,|u_m^j\rangle$$

gives

$$\langle\psi_m|\psi_m\rangle = \left(\sum_{j=1}^{g_m}(c_m^j)^*\,\langle u_m^j|\right)\left(\sum_{k=1}^{g_m} c_m^k\,|u_m^k\rangle\right) = \sum_{j,k=1}^{g_m}(c_m^j)^* c_m^k \delta_{jk} = \sum_{j=1}^{g_m}\left|c_m^j\right|^2 = P(\lambda_m)$$

such that this probability can be written as $P(\lambda_m) = \langle\psi_m|\psi_m\rangle = \langle\psi|\hat{P}_m^\dagger\hat{P}_m|\psi\rangle = \langle\psi|\hat{P}_m^2|\psi\rangle = \langle\psi|\hat{P}_m|\psi\rangle$ using the fact that the projection operator is Hermitian and idempotent. It follows that one can write the probability of a measurement outcome as the expectation value of the projection operator onto the subspace spanned by the eigenstates associated with the measured eigenvalue.

In the case of a continuous eigenvalue spectrum the eigenvalue equation is

$$\hat{A}\,|v_\alpha\rangle = \alpha\,|v_\alpha\rangle$$

where the eigenvalue α is a continuous variable. An arbitrary state $|\psi\rangle$ can be expanded

$$|\psi\rangle = \int d\alpha\, c(\alpha)\,|v_\alpha\rangle \qquad\qquad c(\alpha) = \langle v_\alpha|\psi\rangle$$

as shown above. Probabilities become probability densities, and the probability of a measurement outcome in the range $(\alpha, \alpha + d\alpha)$ is

$$dP(\alpha) = |\langle v_\alpha|\psi\rangle|^2 d\alpha = |c(\alpha)|^2 d\alpha$$

where $|\langle v_\alpha|\psi\rangle|^2 = |c(\alpha)|^2$ is the probability density. A rather famous example of a continuous variable is position which leads to the position representation

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

in one dimension where $\psi(x)$ is the wave function. The probability density of a particle at x is $|\psi(x)|^2$.

1.10 State Collapse

Before a measurement it is impossible to predict what the result will be. Only the likelihood of a result is known. The question is what happens after a measurement. The act of measuring a quantum particle changes its state, and the state after the measurement is an eigenstate corresponding to the measured eigenvalue.

Postulate V: If the measurement of $\mathcal{A}$ in a system in state $|\psi\rangle$ gives eigenvalue λ_m, the state of the system immediately after the measurement is the eigenvalue $|u_m\rangle$ associated with the eigenvalue λ_m that has been measured.

Thus, because of a measurement with outcome λ_m the state collapses from $|\psi\rangle$ to $|u_m\rangle$. This is the case for non-degenerate, discrete eigenvalues. In the case of degenerate eigenvalues one can use the projection operator $P(\lambda_m) = \langle\psi_m|\psi_m\rangle = \langle\psi|\hat{P}_m|\psi\rangle$, and the state is

$$\frac{\hat{P}_m\,|\psi\rangle}{\sqrt{\langle\psi|\hat{P}_m|\psi\rangle}}$$

after the measurement of eigenvalue λ_m. This means that the state immediately after the measurement of λ_m is the normalized projection of $|\psi\rangle$ onto the subspace spanned by the eigenvectors associated with eigenvalue λ_m. In the notations used above $\hat{P}_m\,|\psi\rangle = |\psi_m\rangle$ and $\langle\psi|\hat{P}_m|\psi\rangle = \langle\psi_m|\psi_m\rangle$ this equation shows that the state immediately after the measurement is just $|\psi_m\rangle$ normalized.

In the case of a continuous eigenvalue spectrum the probability is

$$P(\alpha_1 < \alpha < \alpha_2) = \int_{\alpha_1}^{\alpha_2} d\alpha\,|\langle v_\alpha|\psi\rangle|^2 = \int_{\alpha_1}^{\alpha_2} d\alpha\,|c(\alpha)|^2$$

for α in the interval $\Delta = (\alpha_1, \alpha_2)$. Using the projection operator

$$\hat{P}_\Delta = \int_{\alpha_1}^{\alpha_2} d\alpha\ |v_\alpha\rangle\langle v_\alpha|$$

the state $|\psi\rangle$ becomes

$$\frac{\hat{P}_\Delta\,|\psi\rangle}{\sqrt{\langle\psi|\hat{P}_\Delta|\psi\rangle}}$$

immediately after a measurement of a value $\alpha \in \Delta$. For the position operator this is called the collapse of the wave function.

1.11 Expectation Value and Mean Square Deviation

Information captured by a quantum state can be represented by a probability distribution. However, it is important to realize that expectation values and the outcomes of measurements are two completely different concepts in quantum mechanics. If one marks the eigenvalues λ on the x-axis and $P(\lambda)$ on the y-axis one gets a probability distribution. In the following a discrete spectrum is assumed but continuous spectra can be handled very similarly.

In the special case that $|\psi\rangle$ is one of the eigenstates $|u_m\rangle$ the coefficients are $c_m = 1$ and $c_j = 0$ for $j \neq m$, and the probability for measuring λ_m is $P(\lambda_m) = 1$. Thus, one measures λ_m with hundred percent certainty. Here the probability distribution is everywhere zero except for a peak one at λ_m. Since probability distributions can become very complicated there are quantities that allow to simplify (and reduce) the information encoded in a probability distribution.

The first quantity is the *expectation value*. If there are very many copies of the system then quantum mechanics tells the exact fractions of measurements that give a particular outcome. In this context the expectation value is the average value of all these measurement outcomes. The expectation value of the observable $\hat{A}$ is

$$\langle\hat{A}\rangle_\psi = \sum_j \lambda_j P(\lambda_j) = \sum_j \lambda_j |\langle u_j|\psi\rangle|^2$$

given that the system is in state $|\psi\rangle$. If the state is clear from the context the expectation value is usually simply written as $\langle\hat{A}\rangle$ without specifying the state $|\psi\rangle$. However, the expectation value is always related to a state.

The expectation value can be determined differently using

$$\begin{aligned}\langle\psi|\hat{A}|\psi\rangle = \langle\psi|\mathbb{I}\hat{A}\mathbb{I}|\psi\rangle &= \langle\psi|\left(\sum_j |u_j\rangle\langle u_j|\right)\hat{A}\left(\sum_k |u_k\rangle\langle u_k|\right)|\psi\rangle = \sum_{jk}\langle\psi\,|u_j\rangle\langle u_j|\,\hat{A}\,|u_k\rangle\langle u_k|\,\psi\rangle \\ &= \sum_{jk}\lambda_k\,\langle\psi|u_j\rangle\,\langle u_j|u_k\rangle\,\langle u_k|\psi\rangle = \sum_{jk}\lambda_k\,\langle\psi|u_j\rangle\,\delta_{jk}\,\langle u_k|\psi\rangle = \sum_j \lambda_j\,\langle\psi|u_j\rangle\,\langle u_j|\psi\rangle \\ &= \sum_j \lambda_j|\langle u_j|\psi\rangle|^2 = \langle\hat{A}\rangle_\psi\end{aligned}$$

because $\hat{A}\,|u_k\rangle = \lambda_k\,|u_k\rangle$. This leads to the definition of the expectation value

$$\langle\hat{A}\rangle_\psi = \langle\psi|\hat{A}|\psi\rangle \tag{1.9}$$

in its well-known form. One usually works with normalized states $|\psi\rangle$ but the expectation value can be written as

$$\langle\hat{A}\rangle_\psi = \frac{\langle\psi|\hat{A}|\psi\rangle}{\langle\psi|\psi\rangle}$$

in case $|\psi\rangle$ is not normalized.

In the special case $|\psi\rangle = |u_j\rangle$ where $|\psi\rangle$ is an eigenstate the expectation value is

$$\langle\hat{A}\rangle_{u_j} = \langle u_j|\hat{A}|u_j\rangle = \lambda_j\,\langle u_j|u_j\rangle = \lambda_j$$

and is therefore – not very surprisingly – simply the eigenvalue associated with $|u_j\rangle$. In this case the expectation value does coincide with the possible outcome of a measurement, but this is not always the case. If $\hat{A}$ has the two eigenvalues $+1$ and -1 with the corresponding eigenstates $|u_+\rangle$ and $|u_-\rangle$, respectively, then the superposition $|\psi\rangle = \frac{1}{\sqrt{2}}(|u_+\rangle + |u_-\rangle)$ has the expectation value

$$\langle\hat{A}\rangle_\psi = \sum_j \lambda_j P(\lambda_j) = (+1)P(+1) + (-1)P(-1) = \frac{1}{2} - \frac{1}{2} = 0$$

and this is not a possible outcome of a measurement.

The expectation value is a useful quantity but gives only a rough idea about the probability distribution. Any two probability distributions symmetric about the same center lead to the same expectation value, and the expectation value alone does not allow to tell these two probability distributions apart although they can obviously be very different. A quantity describing the width of a probability distribution would therefore help in this situation.

The *mean square deviation* is based on the operator $\hat{\sigma}_A = \hat{A} - \langle\hat{A}\rangle$ and is defined as the expectation value $\langle\hat{\sigma}_A^2\rangle = \langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle$. It can be written as

$$\langle\hat{\sigma}_A^2\rangle = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2 \tag{1.10}$$

because

$$\langle\hat{\sigma}_A^2\rangle = \langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle = \langle\hat{A}^2 - 2\hat{A}\,\langle\hat{A}\rangle + \langle\hat{A}\rangle^2\rangle = \langle\hat{A}^2\rangle - 2\,\langle\hat{A}\rangle^2 + \langle\hat{A}\rangle^2$$

using the fact that the expectation value of a real number is this number such that $\langle\langle\hat{A}\rangle\rangle = \langle\hat{A}\rangle$. The quantity typically used to measure the width of a probability distribution is the square root of the mean square deviation. This ensures that the width has the same units as the quantity itself. Thus, the *root mean square deviation* is defined as

$$\Delta\hat{A} = \sqrt{\langle\hat{\sigma}_A^2\rangle} = \sqrt{\langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2} \tag{1.11}$$

for measuring the width of a probability distribution.

The definition of $\hat{\sigma}_A = \hat{A} - \langle\hat{A}\rangle$ shows that $\hat{\sigma}_A$ is looking at how far one is from the average $\langle\hat{A}\rangle$, and $\langle\hat{\sigma}_A\rangle = 0$ because there are as many terms on one side as on the other. This is the reason why one has to look at the value of the square because the terms on either side cannot cancel. Because the expectation value of $\hat{A}$ is λ_j when $|\psi\rangle$ is an eigenstate $|u_j\rangle$ the root mean square deviation $\Delta\hat{A} = 0$ and this is not a surprise for the width of the corresponding probability distribution.

1.12 The Heisenberg Uncertainty Principle

Any quantum state $|\psi\rangle$ can be represented either as $|\psi\rangle = \sum_j c_j |u_j\rangle$ in the basis of eigenstates $\{|u_j\rangle\}$ of one observable $\hat{A}$ with the eigenvalue equation $\hat{A}|u_j\rangle = \lambda_j |u_j\rangle$ or as $|\psi\rangle = \sum_k d_k |v_k\rangle$ in the basis of eigenstates $\{|v_k\rangle\}$ of another observable $\hat{B}$ with eigenvalue equation $\hat{B}|v_k\rangle = \mu_k |v_k\rangle$. The Heisenberg uncertainty principle is the mathematical relation between the coefficients c_j and the coefficients d_k one can use to represent the same quantum state $|\psi\rangle$.

Formally, the *Heisenberg uncertainty principle* states that

$$\Delta\hat{A}\,\Delta\hat{B} \geq \frac{1}{2}\left|\langle[\hat{A},\hat{B}]\rangle\right| \tag{1.12}$$

using the root mean square deviation(1.11), the expectation value (1.9) and the commutator (1.1). This principle is usually introduced for the two operators position and momentum as $\Delta\hat{x}\,\Delta\hat{p} \geq \frac{1}{2}\hbar$ but it is generally applicable to any two observables $\hat{A}$ and $\hat{B}$.

If the two operators $\hat{A}$ and $\hat{B}$ do not commute then the right side of (1.12) is strictly bigger than zero, and if one draws the probability distribution of $|\psi\rangle$ in the basis of the eigenstates of $\hat{A}$ and in the basis of the eigenstates of $\hat{B}$ then the two graphs are not completely independent. Their width is related such that the narrower the one is the wider must be the other. If the probability distribution for $|\psi'\rangle$ is wider than the one for $|\psi\rangle$ when both are represented in the basis of eigenstates of $\hat{A}$ then the probability distribution for $|\psi\rangle$ must be wider than the one for $|\psi'\rangle$ when both are represented in the basis of eigenstates of $\hat{B}$.

The case where $\hat{A}$ and $\hat{B}$ commute is not so interesting. The Heisenberg uncertainty principle still holds but it only states that $\Delta\hat{A}\,\Delta\hat{B} \geq 0$. There are states with $\Delta\hat{A}$ and $\Delta\hat{B}$ arbitrarily small. This makes sense because two commuting observables can be diagonalized simultaneously such that there are states that are eigenstates for both of them. Since the root mean square deviation vanishes for eigenstates, $\Delta\hat{A}$ and $\Delta\hat{B}$ are both zero.

One can turn around the argument. If two observables do not commute then they cannot be diagonalized simultaneously due to the Heisenberg uncertainty principle because otherwise $\Delta\hat{A} = 0$ and $\Delta\hat{B} = 0$ would be possible. They can therefore not be measured simultaneously.

The fact that the Heisenberg uncertainty principle is called an uncertainty principle is the source of many misrepresentations and misinterpretations. If $|\psi\rangle \approx |u_j\rangle$ and therefore $c_j \approx 1$ when represented in the eigenstate basis of $\hat{A}$ because the probability distribution is very narrow and concentrated around λ_j then $P(\lambda_j) = |c_j|^2 \approx 1$ and the outcome of a measurement of observable $\hat{A}$ would almost certainly be λ_j. Thus in this situation one can predict the outcome of a quantum measurement with high confidence before it has been performed. If $|\psi\rangle$ is represented in the basis of eigenstates of observable $\hat{B}$ with $[\hat{A},\hat{B}] \neq 0$ then the probability distribution must be very wide such that all d_k and therefore also all $P(\mu_k)$ are comparable. This means that $\hat{A}$ has a definite value and there is very little uncertainty before the measurement is performed while, by contrast, $\hat{B}$ can have any value and there is a high degree of uncertainty for the outcome of a measurement before the measurement has been performed. This situation gave the principle its name.

The term "uncertainty" does not mean that one does not know something about the given system because one has not been able to characterize it fully due to some limitations of the measurement apparatus. One knows everything one can know about a quantum system, and it is all described by the state $|\psi\rangle$. The Heisenberg uncertainty principle relates different ways of representing the same state.

The well-known version of this principle relates the position operator $\hat{x}$ and the momentum operator $\hat{p}$ with the commutation relation $[\hat{x},\hat{p}] = i\,\hbar$ gives

$$\Delta\hat{x}\,\Delta\hat{p} \geq \frac{1}{2}|\langle i\,\hbar\rangle| = \frac{1}{2}\hbar$$

when applying (1.12). Position and momentum are continuous variables and a state $|\psi\rangle$ is

$$|\psi\rangle = \int dx\,\psi(x)\,|x\rangle \qquad\qquad |\psi\rangle = \int dp\,\bar{\psi}(p)\,|p\rangle$$

on the left side in the position basis and on the right side in the momentum basis where $\psi(x) = \langle x|\psi\rangle$ and $\bar{\psi}(p) = \langle p|\psi\rangle$ are position and momentum wave functions, respectively. The Heisenberg uncertainty principle says that if the graph of $|\psi(x)|^2$ is very narrow then the graph of $\left|\bar{\psi}(p)\right|^2$ must be very broad and vice versa. The two probability distributions are related by a Fourier transform, and the Fourier transform already encodes the Heisenberg uncertainty principle.

The Heisenberg uncertainty principle can be proven mathematically. It is based on several theorems and lemmata from different areas of mathematics.

The first one is the Schwarz inequality in the form for states in a state space. Given $|\chi\rangle = |\psi\rangle + \alpha\,|\varphi\rangle$ it follows from $\langle\chi|\chi\rangle \geq 0$ that

$$\langle\chi|\chi\rangle = (\langle\psi| + \alpha^*\,\langle\varphi|)\,(|\psi\rangle + \alpha\,|\varphi\rangle) = \langle\psi|\psi\rangle + \alpha\,\langle\psi|\varphi\rangle + \alpha^*\,\langle\varphi|\psi\rangle + |\alpha|^2\,\langle\varphi|\varphi\rangle \geq 0$$

is true for any α. Inserting a special value for α gives

$$\alpha = -\frac{\langle\varphi|\psi\rangle}{\langle\varphi|\varphi\rangle} \in \mathbb{C} \qquad \langle\psi|\psi\rangle - \frac{\langle\varphi|\psi\rangle}{\langle\varphi|\varphi\rangle}\,\langle\psi|\varphi\rangle + \left(-\frac{\langle\varphi|\psi\rangle}{\langle\varphi|\varphi\rangle}\right)^*\langle\varphi|\psi\rangle + \left|\frac{\langle\varphi|\psi\rangle}{\langle\varphi|\varphi\rangle}\right|^2\langle\varphi|\varphi\rangle \geq 0$$

$$\langle\psi|\psi\rangle - \frac{|\langle\psi|\varphi\rangle|^2}{\langle\varphi|\varphi\rangle} - \frac{|\langle\psi|\varphi\rangle|^2}{\langle\varphi|\varphi\rangle} + \frac{|\langle\psi|\varphi\rangle|^2}{\langle\varphi|\varphi\rangle} \geq 0$$

using $\langle\varphi|\psi\rangle^* = \langle\psi|\varphi\rangle$ and $|\langle\psi|\varphi\rangle|^2 = \langle\varphi|\psi\rangle\,\langle\psi|\varphi\rangle$. The result is

$$\langle\psi|\psi\rangle\,\langle\varphi|\varphi\rangle \geq |\langle\psi|\varphi\rangle|^2 \tag{1.13}$$

since two terms cancel.

The second property to be shown is that $\hat{\sigma}_A = \hat{A} - \langle\hat{A}\rangle$ for a Hermitian operator $\hat{A}$ is also Hermitian. This follows from

$$\hat{\sigma}_A^\dagger = \left(\hat{A} - \langle\hat{A}\rangle\right)^\dagger = \hat{A}^\dagger - \langle\hat{A}\rangle^\dagger = \hat{A} - \langle\hat{A}\rangle = \hat{\sigma}_A$$

using just the fact that $\hat{A}$ is Hermitian.

Defining $|\psi_A\rangle = \hat{\sigma}_A\,|\psi\rangle$ and $|\psi_B\rangle = \hat{\sigma}_B\,|\psi\rangle$ leads to

$$\langle\psi_A|\psi_A\rangle\,\langle\psi_B|\psi_B\rangle \geq |\langle\psi_A|\psi_B\rangle|^2$$

using (1.13). The two terms in the Schwarz inequality can be written as

$$\langle\psi_A|\psi_A\rangle = \langle\psi|\hat{\sigma}_A^\dagger\hat{\sigma}_A|\psi\rangle = \langle\psi|\hat{\sigma}_A^2|\psi\rangle = \langle\hat{\sigma}_A^2\rangle \qquad \langle\psi_B|\psi_B\rangle = \langle\psi|\hat{\sigma}_B^\dagger\hat{\sigma}_B|\psi\rangle = \langle\psi|\hat{\sigma}_B^2|\psi\rangle = \langle\hat{\sigma}_B^2\rangle$$

and are therefore the expectation values of $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$, respectively. Similarly, $\langle\psi_A|\psi_B\rangle = \langle\hat{\sigma}_A\hat{\sigma}_B\rangle$ can be shown. This gives

$$\langle\hat{\sigma}_A^2\rangle\,\langle\hat{\sigma}_B^2\rangle \geq |\langle\hat{\sigma}_A\hat{\sigma}_B\rangle|^2$$

inserted into the Schwarz inequality.

A third property needed relates commutator and anticommutator and is

$$\frac{1}{2}[\hat{X},\hat{Y}] + \frac{1}{2}\{\hat{X},\hat{Y}\} = \frac{1}{2}\left(\hat{X}\hat{Y} - \hat{Y}\hat{X}\right) + \frac{1}{2}\left(\hat{X}\hat{Y} + \hat{Y}\hat{X}\right) = \hat{X}\hat{Y}$$

for any two operators $\hat{X}$ and $\hat{Y}$. Applied to $\hat{\sigma}_A$ and $\hat{\sigma}_B$ leads to

$$\hat{\sigma}_A\hat{\sigma}_B = \frac{1}{2}[\hat{\sigma}_A,\hat{\sigma}_B] + \frac{1}{2}\{\hat{\sigma}_A,\hat{\sigma}_B\}$$

and using

$$\begin{aligned}[\hat{\sigma}_A,\hat{\sigma}_B] &= \hat{\sigma}_A\hat{\sigma}_B - \hat{\sigma}_B\hat{\sigma}_A = \left(\hat{A} - \langle\hat{A}\rangle\right)\left(\hat{B} - \langle\hat{B}\rangle\right) - \left(\hat{B} - \langle\hat{B}\rangle\right)\left(\hat{A} - \langle\hat{A}\rangle\right) \\ &= \hat{A}\hat{B} - \hat{A}\,\langle\hat{B}\rangle - \langle\hat{A}\rangle\,\hat{B} + \langle\hat{A}\rangle\,\langle\hat{B}\rangle - \hat{B}\hat{A} + \hat{B}\,\langle\hat{A}\rangle + \langle\hat{B}\rangle\,\hat{A} - \langle\hat{B}\rangle\,\langle\hat{A}\rangle = \hat{A}\hat{B} - \hat{B}\hat{A} = [\hat{A},\hat{B}]\end{aligned}$$

gives

$$\hat{\sigma}_A\hat{\sigma}_B = \frac{1}{2}[\hat{A},\hat{B}] + \frac{1}{2}\{\hat{\sigma}_A,\hat{\sigma}_B\}$$

in this case.

The result so far is

$$\langle\hat{\sigma}_A^2\rangle\,\langle\hat{\sigma}_B^2\rangle \geq |\langle\hat{\sigma}_A\hat{\sigma}_B\rangle|^2 = \left|\frac{1}{2}\,\langle[\hat{A},\hat{B}]\rangle + \frac{1}{2}\,\langle\{\hat{\sigma}_A,\hat{\sigma}_B\}\rangle\right|^2$$

with a commutator and an anticommutator to be determined. Because of

$$[\hat{A},\hat{B}]^\dagger = \left(\hat{A}\hat{B} - \hat{B}\hat{A}\right)^\dagger = \hat{B}^\dagger\hat{A}^\dagger - \hat{A}^\dagger\hat{B}^\dagger = \hat{B}\hat{A} - \hat{A}\hat{B} = -[\hat{A},\hat{B}]$$

the commutator of two Hermitian operators is anti-Hermitian, and the expectation value of an anti-Hermitian operator $\hat{X}^\dagger = -\hat{X}$ is

$$\langle\hat{X}\rangle^* = \langle\psi|\hat{X}|\psi\rangle^* = \langle\psi|\hat{X}^\dagger|\psi\rangle = -\,\langle\psi|\hat{X}|\psi\rangle = -\,\langle\hat{X}\rangle$$

and therefore $\langle[\hat{A},\hat{B}]\rangle$ is purely imaginary. Because of

$$\{\hat{\sigma}_A,\hat{\sigma}_B\}^\dagger = (\hat{\sigma}_A\hat{\sigma}_B + \hat{\sigma}_B\hat{\sigma}_A)^\dagger = \hat{\sigma}_B\hat{\sigma}_A + \hat{\sigma}_A\hat{\sigma}_B = \{\hat{\sigma}_A,\hat{\sigma}_B\}$$

the anticommutator of two Hermitian operators is Hermitian, and the expectation value of a Hermitian opertor $\hat{Y}^\dagger = \hat{Y}$ is

$$\langle\hat{Y}\rangle^* = \langle\psi|\hat{Y}|\psi\rangle^* = \langle\psi|\hat{Y}^\dagger|\psi\rangle = \langle\psi|\hat{Y}|\psi\rangle = \langle\hat{Y}\rangle$$

and therefore $\langle\{\hat{\sigma}_A,\hat{\sigma}_B\}\rangle$ is purely real. Thus,

$$\langle\hat{\sigma}_A^2\rangle\,\langle\hat{\sigma}_B^2\rangle \geq \left|\frac{1}{2}\,\langle[\hat{A},\hat{B}]\rangle\right|^2 + \left|\frac{1}{2}\,\langle\{\hat{\sigma}_A,\hat{\sigma}_B\}\rangle\right|^2$$

because $|x + i\,y|^2 = |x|^2 + |y|^2$. Since both terms on the right side of this inequality are positive, one can drop one of them, and the result with its square root becomes

$$\langle\hat{\sigma}_A^2\rangle\,\langle\hat{\sigma}_B^2\rangle \geq \left|\frac{1}{2}\,\langle[\hat{A},\hat{B}]\rangle\right|^2 \qquad\qquad \sqrt{\langle\hat{\sigma}_A^2\rangle}\sqrt{\langle\hat{\sigma}_B^2\rangle} \geq \frac{1}{2}\left|\langle[\hat{A},\hat{B}]\rangle\right|$$

concluding the proof for (1.12).

1.13 The Schrödinger Equation

The Schrödinger equation is a key equation in quantum physics because it describes the time evolution of quantum states. It therefore corresponds to Newton's second law in classical mechanics.

Postulate VI: The time evolution of a state $|\psi(t)\rangle$ is given by the *Schrödinger equation*

$$i\,\hbar\frac{d}{dt}\,|\psi(t)\rangle = \hat{H}(t)\,|\psi(t)\rangle \tag{1.14}$$

where the Hamilton operator $\hat{H}(t)$ is the observable associated with the total energy of the system.

The Schrödinger equation (1.14) is a postulate and is not something that can be derived. However, all the predictions coming out of this equation are confirmed by experiment. The Planck constant h appears as $\hbar = \frac{h}{2\pi}$ called the reduced Planck constant. The Schrödinger equation states that the time evolution of a quantum system is governed by the total energy of the system.

Time evolution is described in form of a first order differential equation in time t. Thus, given a general solution and initial conditions $|\psi(t)\rangle$ will exactly tell the state of the system at any time t. This means

that the time evolution is deterministic, and there is no ambiguity or probability involved. Probability plays only a role when one performs a measurement.

Combining the Schrödinger equation with measurements provides a more complete picture of time evolution. If the state of the system at time t_0 is $|\psi(t_0)\rangle$ then state at a later time t_1 is $|\psi(t_1)\rangle$ completely determined by the Schrödinger equation as long as no measurement has been performed in the time interval between t_0 and t_1. If a measurement of the observable $\hat{A}$ is performed at time t_1 its outcome is one of the eigenvalues λ_j of $\hat{A}$ with probability $P(\lambda_j) = |\langle u_j|\psi(t_1)\rangle|^2$ where $|u_j\rangle$ is the eigenstate associated with λ_j. After the measurement of λ_k the state changes from $|\psi(t_1)\rangle$ to the eigenstate $|u_k(t_1)\rangle$. Thus, as long as the system is not measured it evolves deterministically according to the Schrödinger equation, but during the measurement it changes probabilistically and instantaneously to the eigenstate corresponding to the measured eigenvalue. After the measurement the state evolves again deterministically according to the Schrödinger equation until another measurement is performed.

One question in this context is what the Hamiltonian of the system is. It consists of the kinetic energy

$$\frac{\hat{p}^2}{2m}$$

in the non-relativistic setting and the potential energy

$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{|\hat{r}_1 - \hat{r}_2|}$$

as an example of the Coulomb force between two electrons. The kinetic energy is always the same but the potential energy depends on the system.

The next question is how one solves the Schrödinger equation. For the conservative systems where the Hamiltonian is time independent such that $\hat{H}(t) = \hat{H}$ there is a general solution. (The solution for time dependent Hamiltonians $\hat{H}(t)$ is more complicated.) A good representation for solving this equation is obviously the energy representation in terms of eigenvalues and eigenstates of $\hat{H}$. The eigenvalue equation for the energy operator is

$$\hat{H}\,|u_j\rangle = E_j\,|u_j\rangle \tag{1.15}$$

in the basis of the energy eigenstates. The Hamiltonian is Hermitian because it is an observable, and any state $|\psi(t)\rangle$ can be written as

$$|\psi(t)\rangle = \sum_j c_j(t)\,|u_j\rangle \qquad\qquad c_j(t) = \langle u_j|\psi(t)\rangle$$

in terms of these time independent eigenstates. The dependence on time is captured in the coefficients $c_j(t)$. The Schrödinger equation becomes

$$\langle u_j|i\,\hbar\frac{d}{dt}|\psi(t)\rangle = \langle u_j|\hat{H}|\psi(t)\rangle \qquad i\,\hbar\frac{d}{dt}\,\langle u_j|\psi(t)\rangle = E_j\,\langle u_j|\psi(t)\rangle \qquad i\,\hbar\frac{d}{dt}c_j(t) = E_j\,c_j(t)$$

in this basis and is therefore a first-order differential equation for the coefficients $c_j(t)$. Integrating gives

$$\int_{t_0}^{t_1}\frac{dc_j(t)}{c_j(t)} = -\frac{i}{\hbar}E_j\int_{t_0}^{t_1}dt \qquad \big[\ln\big(c_j(t)\big)\big]_{t_0}^{t_1} = -\frac{i\,E_j}{\hbar}\,[t]_{t_0}^{t_1} \qquad \ln\left(\frac{c_j(t_1)}{c_j(t_0)}\right) = -\frac{i\,E_j}{\hbar}(t_1 - t_0)$$

and the solution becomes

$$c_j(t) = c_j(t_0)\,e^{-i\,E_j(t-t_0)/\hbar} \qquad\qquad |\psi(t)\rangle = \sum_j c_j(t_0)\,e^{-i\,E_j(t-t_0)/\hbar}\,|u_j\rangle \tag{1.16}$$

for the time dependent coefficients. The solution of the Schrödinger equation for a conservative system is therefore always the same where the coefficients $c_j(t)$ evolve independently in time.

The case where the initial state is an eigenstate of the Hamiltonian is called a *stationary state*. Given $\hat{H}\,|u_j\rangle = E_j\,|u_j\rangle$ and $|\psi(t_0)\rangle = |u_k\rangle$ one can see in

$$|\psi(t_0)\rangle = |u_k\rangle \qquad\qquad |\psi(t)\rangle = e^{-i\,E_k(t-t_0)/\hbar}\,|u_k\rangle$$

that $|\psi(t_0)\rangle$ and $|\psi(t)\rangle$ only differ by a global phase factor. The two states represent the same physical situation. Therefore, a system that is in an eigenstate of the Hamiltonian does not change with time, and this is the reason these states are called stationary.

Conservation of the norm $\langle\psi(t)|\psi(t)\rangle$

$$\begin{aligned}\frac{d}{dt}\langle\psi(t)|\psi(t)\rangle &= \left(\frac{d}{dt}\langle\psi(t)|\right)|\psi(t)\rangle + \langle\psi(t)|\left(\frac{d}{dt}|\psi(t)\rangle\right) \\ &= -\frac{1}{i\,\hbar}\langle\psi(t)|\hat{H}(t)|\psi(t)\rangle + \frac{1}{i\,\hbar}\langle\psi(t)|\hat{H}(t)|\psi(t)\rangle = 0\end{aligned}$$

follows from using the Schrödinger equation (1.14)

$$\frac{d}{dt}|\psi(t)\rangle = \frac{1}{i\,\hbar}\hat{H}(t)\,|\psi(t)\rangle \qquad\qquad \frac{d}{dt}\langle\psi(t)| = -\frac{1}{i\,\hbar}\langle\psi(t)|\,\hat{H}(t)$$

for kets and bras. Thus, one has to normalize a state only once and does not have to worry later about normalization for the state during time evolution.

To summarize, the recipe for conservative systems with a time independent Hamiltonian is:

1. Solve the eigenvalue equation for the Hamiltonian $\hat{H}\,|u_j\rangle = E_j\,|u_j\rangle$ to determine the energy eigenvalues E_j and eigenstates $|u_j\rangle$.
2. Write the initial state of the system as $|\psi(t_0)\rangle = \sum_j c_j(t_0)\,|u_j\rangle$ in the basis of the energy eigenstates.
3. Finally, the state $|\psi(t)\rangle$ at a later time t is given by the equation (1.16).

The case for time dependent Hamiltonian is more difficult.