Introduction to Quantum Physics

Rainer F. Hauser
rainer.hauser@gmail.com

January 14, 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 physics. This script covers the calculus of operators and helps to digest the topic covered by a group of those videos but is not meant as a replacement for them.

2 The Calculus of Operators in Quantum Mechanics

2.1 Unitary Operators

Unitary operators are important in quantum physics as they are used, for example, to describe spatial translations and time evolution. The symmetry operations of a Hamiltonian that can be represented by a group of unitary operators provide another example.

A *unitary operator* $\hat{U}$ is defined by the property $\hat{U}^{-1} = \hat{U}^\dagger$, and there is the alternative definition that an operator $\hat{U}$ is unitary if $\hat{U}^\dagger\hat{U} = \hat{U}\hat{U}^\dagger = \mathbb{I}$ using the definition of the inverse operator $\hat{U}^{-1}\hat{U} = \hat{U}\hat{U}^{-1} = \mathbb{I}$. Unitary operators have several properties.

The product of two unitary operators $\hat{U}$ and $\hat{V}$ is also unitary. This follows from

$$(\hat{U}\hat{V})^\dagger(\hat{U}\hat{V}) = \hat{V}^\dagger\hat{U}^\dagger\hat{U}\hat{V} = \hat{V}^\dagger\mathbb{I}\hat{V} = \hat{V}^\dagger\hat{V} = \mathbb{I}$$

because $(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger$. Similarly, also $(\hat{U}\hat{V})(\hat{U}\hat{V})^\dagger = \mathbb{I}$ can be shown.

The eigenvalue equation for a unitary operator $\hat{U}$ is $\hat{U}\,|\lambda\rangle = \lambda\,|\lambda\rangle$ and the eigenvalues for non-Hermitian operators such as the unitary operators can be complex numbers. Writing the norm in two ways as

$$\left\|\hat{U}\,|\lambda\rangle\right\|^2 = \|\lambda\,|\lambda\rangle\|^2 = \lambda^*\lambda\,\langle\lambda|\lambda\rangle = |\lambda|^2 \qquad \left\|\hat{U}\,|\lambda\rangle\right\|^2 = \langle\lambda|\hat{U}^\dagger\hat{U}|\lambda\rangle = \langle\lambda|\lambda\rangle = 1$$

shows $|\lambda|^2 = 1$. Any complex number with $|\lambda| = 1$ can be written as $\lambda = e^{i\,\varphi_\lambda}$ with $\varphi_\lambda \in \mathbb{R}$. Thus, the eigenvalues of unitary operators can be complex numbers but their magnitude is 1.

From this eigenvalue equation $\hat{U}\,|\lambda\rangle = \lambda\,|\lambda\rangle$ one can conclude $\langle\lambda|\,\hat{U} = \lambda\,\langle\lambda|$. This follows from

$$|\lambda\rangle = \mathbb{I}\,|\lambda\rangle = \hat{U}^\dagger\hat{U}\,|\lambda\rangle = \lambda\,\hat{U}^\dagger\,|\lambda\rangle = e^{i\,\varphi_\lambda}\hat{U}^\dagger\,|\lambda\rangle \qquad \hat{U}^\dagger\,|\lambda\rangle = e^{-i\,\varphi_\lambda}\,|\lambda\rangle = \lambda^*\,|\lambda\rangle$$

and the fact that $\hat{U}^\dagger\,|\lambda\rangle = \lambda^*\,|\lambda\rangle$ is $\langle\lambda|\,\hat{U} = \lambda\,\langle\lambda|$ in the dual space. This property allows to show that eigenstates of unitary operators are orthogonal because from

$$\hat{U}\,|\lambda\rangle = \lambda\,|\lambda\rangle \Rightarrow \langle\mu|\hat{U}|\lambda\rangle = \lambda\,\langle\mu|\lambda\rangle \qquad \langle\mu|\,\hat{U} = \mu\,\langle\mu| \Rightarrow \langle\mu|\hat{U}|\lambda\rangle = \mu\,\langle\mu|\lambda\rangle$$

follows $0 = (\lambda - \mu)\,\langle\mu|\lambda\rangle$ and therefore $\langle\mu|\lambda\rangle = 0$ for $\lambda \neq \mu$. Thus, the eigenstates associated with distinct eigenstates are not only orthogonal for Hermitian operators but also for unitary operators.

Applying a unitary operator to a state or another operator is called *unitary transformation*. A key reason for the importance of unitary operators in quantum mechanics is the fact that they conserve the scalar product as $\langle\psi_1'|\psi_2'\rangle = \langle\psi_1|\hat{U}^\dagger\hat{U}|\psi_2\rangle = \langle\psi_1|\mathbb{I}|\psi_2\rangle = \langle\psi_1|\psi_2\rangle$ shows for $|\psi_1'\rangle = \hat{U}\,|\psi_1\rangle$ and $|\psi_2'\rangle = \hat{U}\,|\psi_2\rangle$, and they therefore also conserve the norm of kets. That means that applying a unitary transformation allows to change the state without changing its normalization. An example is the time evolution of a quantum state which is a process that conserves the norm and can be described by a unitary operator called the *time evolution operator*.

Orthogonal transformations in a real vector space also conserve the scalar product. The rotations or reflections about a point or a plane are examples. Unitary transformations are just the generalization of orthogonal transformation to complex vector spaces.

A unitary transformation of an operator $\hat{A}$ to $\hat{A}'$ is defined by $\langle u_j'|\hat{A}'|u_k'\rangle = \langle u_j|\hat{A}|u_k\rangle$ where the basis $\{|u_j\rangle\}$ is transformed into the basis $\{|u_j'\rangle = \hat{U}\,|u_j\rangle\}$ by a unitary operator $\hat{U}$. From

$$\langle u_j'|\hat{A}'|u_k'\rangle = \langle u_j|\hat{A}|u_k\rangle \qquad\qquad \langle u_j'|\hat{A}'|u_k'\rangle = \langle u_j|\hat{U}^\dagger\hat{A}'\hat{U}|u_k\rangle$$

follows $\hat{A} = \hat{U}^\dagger\hat{A}'\,\hat{U}$, from $\hat{U}\hat{A}\,\hat{U}^\dagger = \hat{U}\hat{U}^\dagger\hat{A}'\hat{U}\hat{U}^\dagger = \mathbb{I}\hat{A}'\mathbb{I} = \hat{A}'$ follows also $\hat{A}' = \hat{U}\hat{A}\,\hat{U}^\dagger$, and the relation $\hat{A}' = \hat{U}\hat{A}\,\hat{U}^\dagger$ between $\hat{A}$ and $\hat{A}'$ expresses the fact that $\hat{A}'$ is the unitary transformation of $\hat{A}$.

If $\hat{A}$ is Hermitian then $\hat{A}'$ is also Hermitian as $(\hat{A}')^\dagger = (\hat{U}\hat{A}\,\hat{U}^\dagger)^\dagger = \hat{U}\hat{A}^\dagger\,\hat{U}^\dagger = (\hat{A}^\dagger)' = \hat{A}'$ shows using $\hat{A}^\dagger = \hat{A}$. Further, $(\hat{A}')^n = (\hat{A}^n)'$ because $(\hat{A}')^n = (\hat{U}\hat{A}\,\hat{U}^\dagger)^n = \hat{U}\hat{A}\,\hat{U}^\dagger\hat{U}\hat{A}\,\hat{U}^\dagger...\hat{U}\hat{A}\,\hat{U}^\dagger = \hat{U}\hat{A}^n\,\hat{U}^\dagger = (\hat{A}^n)'$, and since functions are defined by their power series this implies also $[f(\hat{A})]' = f(\hat{A}')$. Finally, given the eigenvalue equation $\hat{A}\,|\lambda\rangle = \lambda\,|\lambda\rangle$ for $\hat{A}$ the eigenvector equation for $\hat{A}'$ becomes

$$\hat{A}'\,|\lambda'\rangle = \hat{U}\hat{A}\,\hat{U}^\dagger\,\hat{U}\,|\lambda\rangle = \hat{U}\hat{A}\;|\lambda\rangle = \lambda\,\hat{U}\,|\lambda\rangle = \lambda\,|\lambda'\rangle$$

and this means that the eigenvalues stay the same while the eigenstates get transformed.

Any Hermitian operator $\hat{A}^\dagger = \hat{A}$ leads to an unitary operator $\hat{T} = e^{i\,\hat{A}}$ as

$$\hat{T}^\dagger = (e^{i\,\hat{A}})^\dagger = \left(\sum_{k=0}^{\infty}\frac{1}{k!}(i\,\hat{A})^k\right)^\dagger = \sum_{k=0}^{\infty}\frac{1}{k!}(-i\,\hat{A}^\dagger)^k = e^{-i\,\hat{A}^\dagger} = e^{-i\,\hat{A}}$$

shows using the Taylor expansion of $\hat{T}$. Thus, $\hat{T}^\dagger\hat{T} = e^{-i\,\hat{A}}\,e^{i\,\hat{A}} = \mathbb{I} = e^{i\,\hat{A}}\,e^{-i\,\hat{A}} = \hat{T}\hat{T}^\dagger$ as $\hat{A}$ commutes with itself. This implies $\hat{T}^{-1} = \hat{T}^\dagger$. An example is the translation operator $\hat{T}(\alpha) = e^{i\,\alpha\,\hat{p}/\hbar}$ that translates a position eigenstate by an amount α in space. The corresponding Hermitian operator is the momentum operator $\hat{p}$.

An infinitesimal unitary operator is a unitary operator $\hat{U}(\varepsilon)$ that depends on a small number $\varepsilon \in \mathbb{R}$. The statement that $\hat{U}$ is an infinitesimal operator means that $\hat{U}(\varepsilon) \to \mathbb{I}$ as $\varepsilon \to 0$. One can expand $\hat{U}(\varepsilon)$ in a powerseries as $\hat{U}(\varepsilon) = \mathbb{I} + \varepsilon\,\hat{G} + ...$ and $\hat{U}^\dagger(\varepsilon)$ in a powerseries $\hat{U}^\dagger(\varepsilon) = \mathbb{I} + \varepsilon\,\hat{G}^\dagger + ...$ with an operator $\hat{G}$. From $\hat{U}(\varepsilon)\hat{U}^\dagger(\varepsilon) = \mathbb{I} + \varepsilon(\hat{G} + \hat{G}^\dagger) + ... = \mathbb{I}$ follows that $\hat{G} + \hat{G}^\dagger = 0$ or $\hat{G} = -\hat{G}^\dagger$ and this means that $\hat{G}$ is anti-Hermitian. With $\hat{F} = i\hat{G}$ it follows that $\hat{F}^\dagger = \hat{F}$ and that $\hat{F}$ is Hermitian. Therefore, one can always write an infinitesimal unitary operator $\hat{U}(\varepsilon)$ as $\hat{U}(\varepsilon) = \mathbb{I} - i\,\varepsilon\,\hat{F} + ...$ where the operator $\hat{F}$ is Hermitian. An example is the time evolution operator $\hat{U}(t + dt, t) = \mathbb{I} - \frac{i}{\hbar}\hat{H}(t)dt + ...$ where the infinitesimal parameter is dt and the Hermitian operator is the Hamiltonian.

2.2 The Time Evolution Operator

Time evolution of a quantum state is governed by the Schrödinger equation, but the time evolution operator gives an alternative way to see how a system evolves in time. The two alternatives allow to describe time evolution either through states in the Schrödinger picture, through operators in the Heisenberg picture or in a mixed picture called the interaction picture.

The Schrödinger equation is a linear equation such that from $|\psi(t_0)\rangle = c_1\,|\psi_1(t_0)\rangle + c_2\,|\psi_2(t_0)\rangle$ follows that also $|\psi(t)\rangle = c_1\,|\psi_1(t)\rangle + c_2\,|\psi_2(t)\rangle$ at any time t. Thus, the two states $|\psi(t)\rangle$ and $|\psi(t_0)\rangle$ can be related as

$$|\psi(t)\rangle = \hat{U}(t, t_0)\,|\psi(t_0)\rangle \tag{2.1}$$

where $\hat{U}(t,t_0)$ is a linear operator called the *time evolution operator*. This means that applying the time evolution operator to the initial state gives the state at a later time. It has several properties.

One property is that $\hat{U}(t_0,t_0)=\mathbb{I}$ because of $|\psi(t_0)\rangle=\hat{U}(t_0,t_0)\,|\psi(t_0)\rangle$. This makes sense because if the time does not change also the state should not change. For $t\neq t_0$ one can insert equation (2.1) into the Schrödinger equation to get

$$i\,\hbar\frac{d}{dt}\left(\hat{U}(t,t_0)\,|\psi(t_0)\rangle\right)=\hat{H}(t)\,\hat{U}(t,t_0)\,|\psi(t_0)\rangle \qquad\qquad i\,\hbar\frac{d}{dt}\hat{U}(t,t_0)=\hat{H}(t)\,\hat{U}(t,t_0)$$

because $|\psi(t_0)\rangle$ in time independent. This equation can be written differently as

$$\int_{t_0}^{t}\frac{d}{dt'}\hat{U}(t',t_0)\,dt'=\hat{U}(t,t_0)-\hat{U}(t_0,t_0)=-\frac{i}{\hbar}\int_{t_0}^{t}\hat{H}(t')\,\hat{U}(t',t_0)\,dt'$$

by integrating both sides of the equation leading to

$$\hat{U}(t,t_0)=\mathbb{I}-\frac{i}{\hbar}\int_{t_0}^{t}\hat{H}(t')\,\hat{U}(t',t_0)\,dt' \tag{2.2}$$

where the right side may be difficult to solve depending on the system.

The time evolution from t_0 to t can be split into a time evolution from t_0 to t_1 and from t_1 to t as

$$|\psi(t_1)\rangle=\hat{U}(t_1,t_0)\,|\psi(t_0)\rangle \qquad\qquad |\psi(t)\rangle=\hat{U}(t,t_1)\,|\psi(t_1)\rangle$$

such that

$$|\psi(t)\rangle=\hat{U}(t,t_1)\,\hat{U}(t_1,t_0)\,|\psi(t_0)\rangle$$

and this implies

$$\hat{U}(t,t_0)=\hat{U}(t,t_1)\,\hat{U}(t_1,t_0)$$

showing how one can combine time evolution operators together. This result is general and does not require that the times are ordered. Setting $t=t_0$ in this equation gives

$$\hat{U}(t_0,t_0)=\hat{U}(t_0,t_1)\,\hat{U}(t_1,t_0)=\mathbb{I}$$

and one can conclude $\hat{U}(t_0,t_1)=\hat{U}(t_1,t_0)^{-1}$. Thus, going from $\hat{U}(t,t_0)$ to $\hat{U}(t,t_0)^{-1}$ reverses time.

Time evolution is a unitary operator because one can write $d\,|\psi(t)\rangle$ as $|\psi(t+dt)\rangle-|\psi(t)\rangle$ or derive it from the Schrödinger equation to get

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

as an infinitesimal operation and conclude from

$$|\psi(t+dt)\rangle=\left(\mathbb{I}-\frac{i}{\hbar}\hat{H}(t)\,dt\right)|\psi(t)\rangle \qquad\qquad |\psi(t+dt)\rangle=\hat{U}(t+dt,t)\,|\psi(t)\rangle$$

that

$$\hat{U}(t+dt,t)=\mathbb{I}-\frac{i}{\hbar}\hat{H}(t)\,dt \tag{2.3}$$

is the infinitesimal time evolution operator. As shown above $\hat{U}(\varepsilon)=\mathbb{I}-i\,\varepsilon\,\hat{F}$ is unitary if $\hat{F}^\dagger=\hat{F}$, and therefore $\hat{U}(t+dt,t)$ is an infinitesimal unitary operator. Combining many infinitesimal time evolution operators $\hat{U}(t+dt,t)$ gives the general time evolution operator $\hat{U}(t,t')$ and this is unitary as well because the product of unitary operators is unitary. Thus, $\hat{U}(t,t')^\dagger=\hat{U}(t,t')^{-1}=\hat{U}(t',t)$. Because the Schrödinger equation conserves the norm one could have therefore guessed that the time evolution operator must be unitary.

A conservative system is characterized by a time independent Hamiltonian $\hat{H}(t) = \hat{H}$ such that

$$i\,\hbar \frac{d}{dt}\hat{U}(t,t_0) = \hat{H}\,\hat{U}(t,t_0) \qquad\qquad \hat{U}(t,t_0) = e^{-i\,\hat{H}(t-t_0)/\hbar} \tag{2.4}$$

describes the time evolution. Time evolution in conservative systems have a simple form in the energy basis. The eigenvalue equation is $\hat{H}\,|u_j\rangle = E_j\,|u_j\rangle$. This gives

$$|\psi(t)\rangle = \hat{U}(t,t_0)\,|\psi(t_0)\rangle = e^{-i\,\hat{H}(t-t_0)/\hbar}\sum_j c_j\,|u_j\rangle = \sum_j c_j\,e^{-i\,\hat{H}(t-t_0)/\hbar}\,|u_j\rangle = \sum_j c_j\,e^{-i\,E_j(t-t_0)/\hbar}\,|u_j\rangle$$

using the expansion of $|\psi(t_0)\rangle$ in the energy basis. The same result has already been shown using the Schrödinger equation, and this shows that the same result can be obtained using the time evolution operator.

One might guess that the time evolution operators is

$$\hat{U}(t,t_0) \stackrel{?}{=} e^{-\frac{i}{\hbar}\int_{t_0}^{t}\hat{H}(t')\,dt'}$$

for the time dependent Hamiltonian $\hat{H}(t)$. This would be correct if $[\hat{H}(t),\hat{H}(t_0)] = 0$ for $t \neq t_0$ but in the general case one findes $[\hat{H}(t),\hat{H}(t_0)] \neq 0$ for $t \neq t_0$. In this case, the solution can be written as a series called Dyson series.

2.3 Schrödinger and Heisenberg Picture

The Schrödinger equation describes how quantum states evolve in time as determined by the Hamiltonian. This is called the Schrödinger picture. However, time evolution in quantum mechanics can be described in many different ways. In the Heisenberg picture states do not change in time, but it is observables that encode time evolution.

The quantum state at t_0 is $|\psi_{\mathrm{S}}(t_0)\rangle$ and the state at a later time t is $|\psi_{\mathrm{S}}(t)\rangle$. The question is how to go from the state at t_0 to the state at t. The answer so far is given by postulate VI with the Schrödinger equation. This means that time evolution is given by

$$i\,\hbar\frac{d}{dt}\,|\psi_{\mathrm{S}}(t)\rangle = \hat{H}_{\mathrm{S}}(t)\,|\psi_{\mathrm{S}}(t)\rangle$$

in the Schrödinger picture indicated by the index $_{\mathrm{S}}$. Observables such as the momentum operator $\hat{p}_{\mathrm{S}}$ are represented by a Hermitian operator, and these operators are time independent if the observable is time independent. However, observables can be time dependent such as a Hamiltonian $\hat{H}_{\mathrm{S}}(t)$ for a particle in a time dependent potential. Time evolution can also equivalently be described as

$$|\psi_{\mathrm{S}}(t)\rangle = \hat{U}(t,t_0)\,|\psi_{\mathrm{S}}(t_0)\rangle$$

using the unitary time evolution operator (2.1) with the solution (2.4) in case of a conservative system.

Time evolution in quantum mechanics can also be described by many other means. One is the Heisenberg picture. Given an observable $\hat{A}_{\mathrm{S}}$, the expectation value $\langle\psi_{\mathrm{S}}(t)|\hat{A}_{\mathrm{S}}|\psi_{\mathrm{S}}(t)\rangle$ can be seen in two ways as

$$\begin{aligned}\langle\psi_{\mathrm{S}}(t)|\hat{A}_{\mathrm{S}}|\psi_{\mathrm{S}}(t)\rangle &= \langle\psi_{\mathrm{S}}(t_0)|\underline{\hat{U}^\dagger(t,t_0)}\,\hat{A}_{\mathrm{S}}\,\underline{\hat{U}(t,t_0)}|\psi_{\mathrm{S}}(t_0)\rangle \\ &= \langle\psi_{\mathrm{S}}(t_0)|\underline{\hat{U}^\dagger(t,t_0)\,\hat{A}_{\mathrm{S}}\,\hat{U}(t,t_0)}|\psi_{\mathrm{S}}(t_0)\rangle = \langle\psi_{\mathrm{H}}|\hat{A}_{\mathrm{H}}(t)|\psi_{\mathrm{H}}\rangle\end{aligned}$$

with

$$\begin{aligned}&\langle\psi_{\mathrm{S}}(t)| = \langle\psi_{\mathrm{S}}(t_0)|\,\hat{U}^\dagger(t,t_0) \qquad |\psi_{\mathrm{S}}(t)\rangle = \hat{U}(t,t_0)\,|\psi_{\mathrm{S}}(t_0)\rangle \\ &\hat{A}_{\mathrm{H}}(t) = \hat{U}^\dagger(t,t_0)\,\hat{A}_{\mathrm{S}}\,\hat{U}(t,t_0) \quad |\psi_{\mathrm{H}}\rangle = |\psi_{\mathrm{S}}\rangle\,(t_0)\end{aligned} \tag{2.5}$$

with the time dependent state $|\psi_{\mathrm{S}}(t)\rangle$ in the Schrödinger picture and the time dependent observable $\hat{A}_{\mathrm{H}}(t)$ in the Heisenberg picture with the time independent state $|\psi_{\mathrm{S}}\rangle\,(t_0)$. Conceptually, one picks a reference time t_0, looks at what the state is at that time in the Schrödinger picture and takes it as the state in the Heisenberg picture.

Unitary transformations $\hat{U}$ conserve the scalar product of states because

$$\left.\begin{array}{ll} |\psi\rangle \Rightarrow & |\psi'\rangle = \hat{U}\,|\psi\rangle \\ |\varphi\rangle \Rightarrow & |\varphi'\rangle = \hat{U}\,|\varphi\rangle \end{array}\right\} \langle\psi'|\varphi'\rangle = \langle\psi|\varphi\rangle$$

as unitary transformations satisfy $\hat{U}^\dagger\hat{U} = \hat{U}\hat{U}^\dagger = \mathbb{I}$. Unitary transformations therefore also conserve the norm. Unitary transformations correspond in the complex vector spaces to the orthogonal transformations in the real vector spaces. This is an important property because all predictions of quantum mechanics are given as scalar products. Thus, any unitary transformation defines a possible picture of quantum mechanics. In the Heisenberg picture, the unitary operator is the time evolution operator.

The dynamics in the Schrödinger picture is given by the Schrödinger equation

$$i\,\hbar\frac{d}{dt}\,|\psi(t)\rangle = \hat{H}(t)\,|\psi(t)\rangle \tag{2.6}$$

and by

$$i\hbar\frac{d}{dt}\hat{A}_{\mathrm{H}}(t) = \left[\hat{A}_{\mathrm{H}}(t), \hat{H}_{\mathrm{H}}(t)\right] + i\hbar\left[\frac{d}{dt}\hat{A}_{\mathrm{S}}(t)\right]_{\mathrm{H}} \tag{2.7}$$

in the Heisenberg pictures as can be shown using the chain rule and the resolution of the identity in the form $\hat{U}(t,t_0)\hat{U}^\dagger(t,t_0) = \mathbb{I}$ with (2.5) inserted into the Schrödinger equation.

2.4 Commutator Algebra

Operators do not always commute, and the commutator becomes a key quantity in quantum mechanics. Commutators have useful properties needed throughout quantum physics.

The first property of commutators is $[\hat{A},\hat{B}] = -[\hat{B},\hat{A}]$. This follows from

$$[\hat{A},\hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = -(\hat{B}\hat{A} - \hat{A}\hat{B}) = -[\hat{B},\hat{A}]$$

easily. Also the second property $[\hat{A},\hat{B}+\hat{C}] = [\hat{A},\hat{B}] + [\hat{A},\hat{C}]$ can be proven by

$$\begin{aligned}[\hat{A},\hat{B}+\hat{C}] &= \hat{A}(\hat{B}+\hat{C}) - (\hat{B}+\hat{C})\hat{A} = \hat{A}\hat{B} + \hat{A}\hat{C} - \hat{B}\hat{A} - \hat{C}\hat{A} = (\hat{A}\hat{B} - \hat{B}\hat{A}) + (\hat{A}\hat{C} - \hat{C}\hat{A}) \\ &= [\hat{A},\hat{B}] + [\hat{A},\hat{C}]\end{aligned}$$

directly from the definition of a commutator. The third property $[\hat{A},\hat{B}\hat{C}] = \hat{B}[\hat{A},\hat{C}] + [\hat{A},\hat{B}]\hat{C}$ follows from

$$\begin{aligned}[\hat{A},\hat{B}\hat{C}] &= \hat{A}\hat{B}\hat{C} - \hat{B}\hat{C}\hat{A} = \hat{A}\hat{B}\hat{C} - \hat{B}\hat{A}\hat{C} + \hat{B}\hat{A}\hat{C} - \hat{B}\hat{C}\hat{A} = (\hat{A}\hat{B} - \hat{B}\hat{A})\hat{C} + \hat{B}(\hat{A}\hat{C} - \hat{C}\hat{A}) \\ &= [\hat{A},\hat{B}]\hat{C} + \hat{B}[\hat{A},\hat{C}]\end{aligned}$$

by adding and subtracting the same term. The next property $[\hat{A}\hat{B},\hat{C}] = \hat{A}[\hat{B},\hat{C}] + [\hat{A},\hat{C}]\hat{B}$ can be shown similarly. The last property $[\hat{A},\hat{B}]^\dagger = [\hat{B}^\dagger,\hat{A}^\dagger]$ is proven by

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

using $(\hat{X}\hat{Y})^\dagger = \hat{Y}^\dagger\hat{X}^\dagger$. There are other properties such as $[\hat{A},[\hat{B},\hat{C}] + [\hat{B},[\hat{C},\hat{A}] + [\hat{C},[\hat{A},\hat{B}] = 0$ to be proven similarly.

2.5 Functions of Operators

Operators act on states and give other states, and in quantum mechanics operators are linear. Operators often are functions of other operators such as the kinetic energy $\frac{\hat{p}^2}{2m}$ containing the square of the momentum operator $\hat{p}$, and this is a power function. The time evolution operator and the translation operator are exponential functions of the Hamiltonian and the momentum operator, respectively.

The power of an operator is defined as $\hat{A}^n = \hat{A}\hat{A}...\hat{A}$ applied n times with $\hat{A}^0 = \mathbb{I}$. A function $F(\hat{A})$ of an operator can be expanded similar to a function of a scalar variable as

$$F(\hat{A}) = \sum_{n=0}^{\infty} f_n \, \hat{A}^n \tag{2.8}$$

in terms of its powers where the quantities f_n are the expansion coefficients. This is all one needs to know about functions of operators. However, with infinite sums there is always the question of convergence but this question is ignored here. An example of a function of operators is the exponential function with the Taylor expansion

$$e^{\hat{A}} = \sum_{n=0}^{\infty} \frac{1}{n!} \, \hat{A}^n = \mathbb{I} + \hat{A} + \frac{1}{2}\hat{A}^2 + \frac{1}{6}\hat{A}^3 + ...$$

already used in the discussion above about the time evolution operator for a conservative system.

Because scalars always commute but operators not necessarily do so, there are some pitfalls when comparing Taylor expansions of scalar functions with those of functions of operators. For numbers $e^a \, e^b = e^{a+b}$ is always the case but this may not be true for operators as

$$\begin{aligned}
e^{\hat{A}} \, e^{\hat{B}} &= \left(\sum_{m=0}^{\infty} \frac{1}{m!} \, \hat{A}^m \right) \left(\sum_{n=0}^{\infty} \frac{1}{n!} \, \hat{B}^n \right) = \left(\mathbb{I} + \hat{A} + \frac{1}{2}\hat{A}^2 + \frac{1}{6}\hat{A}^3 + ... \right) \left(\mathbb{I} + \hat{B} + \frac{1}{2}\hat{B}^2 + \frac{1}{6}\hat{B}^3 + ... \right) \\
&= \mathbb{I} + \big(\hat{A} + \hat{B}\big) + \frac{1}{2}\big(\hat{A}^2 + \hat{B}^2 + 2\hat{A}\hat{B}\big) + ... \\
e^{\hat{A}+\hat{B}} &= \sum_{n=0}^{\infty} \frac{1}{n!} \, (\hat{A} + \hat{B})^n = \mathbb{I} + \big(\hat{A} + \hat{B}\big) + \frac{1}{2}\big(\hat{A} + \hat{B}\big)^2 + ... \\
&= \mathbb{I} + \big(\hat{A} + \hat{B}\big) + \frac{1}{2}\big(\hat{A}^2 + \hat{B}^2 + \hat{A}\hat{B} + \hat{B}\hat{A}\big) + ...
\end{aligned}$$

shows because $2\hat{A}\hat{B} \neq \hat{A}\hat{B} + \hat{B}\hat{A}$ if $\hat{A}$ and $\hat{B}$ do not commute. The Baker-Campbell-Hausdorff formula

$$e^{\hat{A}} \, e^{\hat{B}} = e^{\hat{C}} \qquad\qquad \hat{C} = \hat{A} + \hat{B} + \frac{1}{2}[\hat{A}, \hat{B}] + \frac{1}{12}\big([\hat{A}, [\hat{A}, \hat{B}]] - [\hat{B}, [\hat{A}, \hat{B}]]\big) + ...$$

is given here without proof. It shows that $e^{\hat{A}} \, e^{\hat{B}} = e^{\hat{A}+\hat{B}}$ if $\hat{A}$ and $\hat{B}$ commute. This is just one example where one has to be cautious when working with functions of operators.

If $\hat{A}$ is Hermitian then

$$\big(F(\hat{A})\big)^\dagger = \left(\sum_{n=0}^{\infty} f_n \, \hat{A}^n \right)^\dagger = \sum_{n=0}^{\infty} f_n^* \, \big(\hat{A}^n\big)^\dagger = \sum_{n=0}^{\infty} f_n^* \, \big(\hat{A}^\dagger\big)^n = \sum_{n=0}^{\infty} f_n^* \, \hat{A}^n$$

using $\hat{A}^\dagger = \hat{A}$. Thus, a function of a Hermitian operator is in the most general case not Hermitian, but if F is a real function then all expansion coefficients are real numbers $f_n \in \mathbb{R}$ and the function $F(\hat{A})$ is Hermitian.

With the eigenvalue equation $\hat{A} \, |u_j\rangle = \lambda_j \, |u_j\rangle$ the eigenvalue equation for $F(\hat{A})$ is

$$F(\hat{A}) \, |u_j\rangle = \sum_{n=0}^{\infty} f_n \, \hat{A}^n \, |u_j\rangle = \sum_{n=0}^{\infty} f_n \, \lambda_j^n \, |u_j\rangle = F(\lambda_j) \, |u_j\rangle$$

and this means that $F(\hat{A})$ for any function F has the same eigenstates as $\hat{A}$ and the eigenvalues are the function F applied to the eigenvalues of $\hat{A}$.

In order to determine the commutator of functions of operators the power functions have to be examined first. The case

$$[\hat{A}, \hat{A}^n] = \hat{A}\hat{A}^n - \hat{A}^n\hat{A} = \hat{A}^{n+1} - \hat{A}^{n+1} = 0$$

is simple. Therefore, also $[\hat{A}, F(\hat{A})] = 0$. Further, from $[\hat{A}, \hat{B}] = 0$ follows

$$[\hat{A}, \hat{B}^n] = \hat{A}\hat{B}^n - \hat{B}^n\hat{A} = 0$$

and also $[\hat{A}, F(\hat{B})] = 0$ for any function F. The general case $[\hat{A}, \hat{B}] \neq 0$ does not lead to nice expressions for the powers, but the special case $[\hat{A}, [\hat{A}, \hat{B}]] = 0$ and $[\hat{B}, [\hat{A}, \hat{B}]] = 0$ is interesting because it is true for the position and momentum operators where $[\hat{x}, \hat{p}] = i\,\hbar$ is a number. In this case $[\hat{A}, \hat{B}^n] = n[\hat{A}, \hat{B}]\hat{B}^{n-1}$ as the proof by induction with the initialization

$$\begin{aligned} n &= 0 & [\hat{A}, \hat{B}^0] &= [\hat{A}, \mathbb{I}] = 0 \\ n &= 1 & [\hat{A}, \hat{B}] &= 1[\hat{A}, \hat{B}]\hat{B}^0 \end{aligned}$$

and the induction step

$$\begin{aligned} [\hat{A}, \hat{B}^{n+1}] &= [\hat{A}, \hat{B}\hat{B}^n] = \hat{B}[\hat{A}, \hat{B}^n] + [\hat{A}, \hat{B}]\hat{B}^n = \hat{B}\,n[\hat{A}, \hat{B}]\hat{B}p^{n-1} + [\hat{A}, \hat{B}]\hat{B}^n = n[\hat{A}, \hat{B}]\hat{B}^n + [\hat{A}, \hat{B}]\hat{B}^n \\ &= (n+1)[\hat{A}, \hat{B}]\hat{B}^n \end{aligned}$$

shows using $[\hat{X}, \hat{Y}\,\hat{Z}] = \hat{Y}[\hat{X}, \hat{Z}] + [\hat{X}, \hat{Y}]\hat{Z}$. One can now apply this to $[\hat{A}, F(\hat{B})]$ and gets

$$\begin{aligned} [\hat{A}, F(\hat{B})] &= [\hat{A}, \sum_{n=0}^{\infty} f_n\,\hat{B}^n] = \sum_{n=0}^{\infty} f_n[\hat{A}, \hat{B}^n] = \sum_{n=0}^{\infty} f_n\,n[\hat{A}, \hat{B}]\hat{B}^{n-1} = [\hat{A}, \hat{B}]\sum_{n=0}^{\infty} n\,f_n\,\hat{B}^{n-1} \\ &= [\hat{A}, \hat{B}]\,F'(\hat{B}) \end{aligned}$$

where $F'(\hat{B})$ is the derivative. One can therefore write $[\hat{A}, F(\hat{B})] = [\hat{A}, \hat{B}]\,F'(\hat{B})$. Applied to the position and the momentum operator gives $[\hat{x}, F(\hat{p})] = i\,\hbar\,F'(\hat{p})$ and $[\hat{p}, G(\hat{x})] = -i\,\hbar\,G'(\hat{x})$ using $[\hat{x}, \hat{p}] = i\,\hbar$.

2.6 Translation Operators

The position operator $\hat{x}$ and the momentum operator $\hat{p}$ do not commute as the well-known case of the Heisenberg uncertainty principle shows, and their commutation relations are $[\hat{x}, \hat{p}] = i\,\hbar$. The *translation operator* is defined as

$$\hat{T}(\alpha) = e^{-i\,\alpha\,\hat{p}/\hbar} \qquad\qquad \alpha \in \mathbb{R} \tag{2.9}$$

and it translates in space by an amount α. The adjoint operator is

$$\hat{T}^\dagger(\alpha) = e^{i\,\alpha\,\hat{p}^\dagger/\hbar} = e^{i\,\alpha\,\hat{p}/\hbar} = e^{-i\,(-\alpha)\,\hat{p}/\hbar} = \hat{T}(-\alpha)$$

using $\hat{p}^\dagger = \hat{p}$. The translation operator is therefore not Hermitian but is unitary as

$$\hat{T}^\dagger(\alpha)\hat{T}(\alpha) = e^{i\,\alpha\,\hat{p}/\hbar}\,e^{-i\,\alpha\,\hat{p}/\hbar} = \mathbb{I} \qquad\qquad \hat{T}(\alpha)\hat{T}^\dagger(\alpha) = e^{-i\,\alpha\,\hat{p}/\hbar}\,e^{i\,\alpha\,\hat{p}/\hbar} = \mathbb{I}$$

shows. Because $[\hat{p}, \hat{p}] = 0$ one can calculate these exponents as if they were just numbers. The translation operator satisfies therefore $\hat{T}^\dagger(\alpha) = \hat{T}^{-1}(\alpha) = \hat{T}(-\alpha)$.

The commutator of the translation operator with the position operator is

$$\left[\hat{x}, \hat{T}(\alpha)\right] = \left[\hat{x}, e^{-i\,\alpha\,\hat{p}/\hbar}\right] = [\hat{x}, \hat{p}]\left(-\frac{i\,\alpha}{\hbar}\right)e^{-i\,\alpha\,\hat{p}/\hbar} = i\,\hbar\left(-\frac{i\,\alpha}{\hbar}\right)\hat{T}(\alpha) = \alpha\,\hat{T}(\alpha)$$

using $[\hat{x}, F(\hat{p})] = [\hat{x}, \hat{p}]\,F'(\hat{p}) = i\,\hbar\,F'(\hat{p})$. This gives all that is needed to proof that the $\hat{T}(\alpha)$ is indeed the translation operator.

Using $\hat{x}\,|x\rangle = x\,|x\rangle$ and $\hat{x}\,\hat{T}(\alpha) = [\hat{x}, \hat{T}(\alpha)] + \hat{T}(\alpha)\hat{x} = \alpha\,\hat{T}(\alpha) + \hat{T}(\alpha)\hat{x}$ one can show that

$$\begin{aligned} \hat{x}\big(\hat{T}(\alpha)\,|x\rangle\big) &= \big(\alpha\,\hat{T}(\alpha) + \hat{T}(\alpha)\hat{x}\big)\,|x\rangle = \alpha\,\hat{T}(\alpha)\,|x\rangle + \hat{T}(\alpha)\hat{x}\,|x\rangle = \alpha\,\hat{T}(\alpha)\,|x\rangle + x\,\hat{T}(\alpha)\,|x\rangle \\ &= (\alpha + x)\big(\hat{T}(\alpha)\,|x\rangle\big) \end{aligned}$$

and this means that if $|x\rangle$ is an eigenstate of $\hat{x}$ with eigenvalue x then $\hat{T}(\alpha)\,|x\rangle$ is an eigenstate of $\hat{x}$ with eigenvalue $(x+\alpha)$ and that $\hat{T}(\alpha)\,|x\rangle = |x+\alpha\rangle$. Thus, $\hat{T}(\alpha)$ is a translation operator that translates a ket $|x\rangle$ by an amount of α. It is also comprehensible that $\hat{T}^{-1}(\alpha) = \hat{T}(-\alpha)$. The translation operator $\hat{T}(-\alpha)\,|x\rangle = |x-\alpha\rangle$ is

$$\langle x-\alpha| = \langle x|\,\hat{T}^{\dagger}(-\alpha) = \langle x|\,\hat{T}(\alpha)$$

in dual space.

The Taylor expansion of the translation of an infinitesimal amount ε is

$$\hat{T}(-\varepsilon) = e^{i\,\varepsilon\,\hat{p}/\hbar} = \mathbb{I} + \frac{i\,\varepsilon}{\hbar}\hat{p} + O(\varepsilon^2)$$

up to order ε. This approximation will turn out to be very useful.

2.7 Changing Basis of States and Observables

For any calculation in quantum mechanics choosing a good basis is important because it simplifies the mathematics. Therefore, changing from one basis to another is essential. To represent kets, bras and operators in the state space a basis $\{|u_j\rangle\}$ is needed and the basis is assumed to be orthonormal such that $\langle u_j|u_k\rangle = \delta_{jk}$. Any state can be represented in this basis as

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

where the complex numbers c_j are called the coefficients of $|\psi\rangle$ in the $\{|u_j\rangle\}$ basis. The corresponding bra in the dual space is

$$\langle\psi| = \sum_j c_j^*\,\langle u_j| \qquad\qquad c_j^* = \langle\psi|u_j\rangle$$

and an operator is

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

represented in the same basis.

In another orthonormal basis $\{|v_j\rangle\}$ the same state $|\psi\rangle$ is $|\psi\rangle = \sum_k d_k\,|v_k\rangle$ with

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

where $S_{kj} = \langle v_k|u_j\rangle$ is called the overlap matrix. To go from $\{|v_j\rangle\}$ to $\{|u_j\rangle\}$ the overlap matrix is $\langle u_j|v_k\rangle = \langle v_k|u_j\rangle^* = S_{kj}^*$. For observables a similar calculation leads to

$$A^v_{km} = \langle v_k|\hat{A}|v_m\rangle = \langle v_k|\mathbb{I}\hat{A}\mathbb{I}|v_m\rangle = \sum_{j\ell} S_{kj}A^u_{j\ell}S^*_{m\ell}$$

in terms of the overlap matrix.

2.8 The Ehrenfest Theorem

Time evolution of the expectation values of operators is the general topic here and Ehrenfest's theorem is a special case where the operator is the position operator or the momentum operator. This theorem shows that the time evolution of the expectation value of position and momentum is described by a set of equations that resemble those obeyed by a classical particle. This allows a direct connection between the classical and the quantum worlds.

The ingredients for the expectation value are a state $|\psi\rangle$ and an observable $\hat{A}$ where the expectation value of $\hat{A}$ in state $|\psi\rangle$ is $\langle\hat{A}\rangle_\psi = \langle\psi|\hat{A}|\psi\rangle$. Since the topic here is the time dependent expectation value, the state is $|\psi(t)\rangle$ and also the observable may depend on time as the Hamiltonian with a time dependent potential does. The expectation value becomes $\langle\hat{A}\rangle(t) = \langle\psi(t)|\hat{A}(t)|\psi(t)\rangle$ where the subscript ψ has been omitted in $\langle\hat{A}\rangle(t)$. The derivative with respect to time t is

$$\begin{aligned}\frac{d}{dt}\left[\langle\hat{A}\rangle(t)\right] &= \frac{d}{dt}\left[\langle\psi(t)|\hat{A}(t)|\psi(t)\rangle\right] \\ &= \left[\frac{d}{dt}\langle\psi(t)|\right]\hat{A}(t)\,|\psi(t)\rangle + \langle\psi(t)|\,\hat{A}(t)\left[\frac{d}{dt}|\psi(t)\rangle\right] + \langle\psi(t)|\frac{\partial}{\partial t}\hat{A}(t)|\psi(t)\rangle\end{aligned}$$

and the Schrödinger equation gives

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

for the ket and the bra. Combining this together results in

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

and in

$$\frac{d}{dt}\langle\hat{A}(t)\rangle = \frac{1}{i\hbar}\left\langle\left[\hat{A}(t),\hat{H}(t)\right]\right\rangle + \left\langle\frac{\partial}{\partial t}\hat{A}(t)\right\rangle \tag{2.10}$$

for the expectation values. This result is general and applies to any observable $\hat{A}$.

If $\hat{A}$ is either the position operator $\underline{\hat{r}}$ or the momentum operator $\underline{\hat{p}}$ then the equation (2.10) reduces to

$$\frac{d}{dt}\langle\hat{A}\rangle = \frac{1}{i\hbar}\left\langle\left[\hat{A},\hat{H}(t)\right]\right\rangle$$

because both observables $\underline{\hat{r}}$ and $\underline{\hat{p}}$ are time independent. For simplicity the system is chosen to be a particle moving in a scalar potential with

$$\hat{H}(t) = \frac{\underline{\hat{p}}^2}{2m} + V(\underline{\hat{r}})$$

as its Hamiltonian. Keeping in mind that $\left[\underline{\hat{r}},\hat{H}(t)\right]$ actially consists of three individual equations this formula becomes for the position operator

$$\frac{d}{dt}\langle\underline{\hat{r}}\rangle = \frac{1}{i\hbar}\left\langle\left[\underline{\hat{r}},\hat{H}(t)\right]\right\rangle = \frac{1}{i\hbar}\left\langle\left[\underline{\hat{r}},\frac{\underline{\hat{p}}^2}{2m}\right] + \left[\underline{\hat{r}},V(\underline{\hat{r}}\right]\right\rangle = \frac{1}{i\hbar}\left\langle\left[\underline{\hat{r}},\frac{\underline{\hat{p}}^2}{2m}\right]\right\rangle = \frac{1}{m}\,\langle\underline{\hat{p}}\rangle$$

using the fact that the position operator commutes with any function of position together with

$$[\hat{x},\underline{\hat{p}}^2] = [\hat{x},\hat{p}_x^2+\hat{p}_y^2+\hat{p}_z^2] = [\hat{x},\hat{p}_x^2] = [\hat{x},\hat{p}_x]\hat{p}_x + \hat{p}_x[\hat{x},\hat{p}_x] = 2\,i\,\hbar\,\hat{p}_x \qquad\qquad \left[\underline{\hat{r}},\frac{\underline{\hat{p}}^2}{2m}\right] = \frac{i\,\hbar}{m}\,\underline{\hat{p}}$$

following from $[\hat{A},\hat{B}\hat{C}] = \hat{B}[\hat{A},\hat{C}] + [\hat{A},\hat{B}]\hat{C}$.

A similar calculation for $\underline{\hat{p}}$ gives

$$\frac{d}{dt}\langle\underline{\hat{p}}\rangle = \frac{1}{i\hbar}\left\langle\left[\underline{\hat{p}},\hat{H}(t)\right]\right\rangle = \frac{1}{i\hbar}\left\langle\left[\underline{\hat{p}},\frac{\underline{\hat{p}}^2}{2m}\right] + \left[\underline{\hat{p}},V(\underline{\hat{r}})\right]\right\rangle = -\,\langle\nabla V(\underline{\hat{r}})\rangle$$

using $[\hat{p}_x,F(\hat{x})] = -i\hbar\,F'(\hat{x})$. Thus, the Ehrenfest theorem is

$$\frac{d}{dt}\langle\underline{\hat{r}}\rangle = \frac{1}{m}\,\langle\underline{\hat{p}}\rangle \qquad\qquad \frac{d}{dt}\langle\underline{\hat{p}}\rangle = -\,\langle\nabla V(\underline{\hat{r}})\rangle \qquad\qquad m\frac{d^2}{dt^2}\langle\underline{\hat{r}}\rangle = -\,\langle\nabla V(\underline{\hat{r}})\rangle \tag{2.11}$$

for position and momentum where the third formula comes from

$$m\frac{d^2}{dt^2}\langle\hat{\underline{r}}\rangle = m\frac{d}{dt}\left(\frac{d}{dt}\langle\hat{\underline{r}}\rangle\right) = m\frac{d}{dt}\left(\frac{1}{m}\langle\hat{\underline{p}}\rangle\right) = \frac{d}{dt}\langle\hat{\underline{p}}\rangle = -\langle\nabla V(\hat{\underline{r}})\rangle$$

using the other two formulas in (2.11). This shows a connection between classical mechanics and quantum mechanics. Newton's second law states

$$m\frac{d^2}{dt^2}\underline{r} = \underline{F} = -\nabla V(\underline{r})$$

in case the force $\underline{F}$ comes from a potential $V(\underline{r})$, and has the same form as the third equation in Ehrenfest's theorem (2.11).

The question is what this connection exactly is. In the position representation is $\psi(\underline{r},t)$ a wave function, and to make it concrete it is assumed that it is a wave packet. The expectation value of the wave packet is at $\langle\hat{\underline{r}}\rangle(t_0)$ and at $\langle\hat{\underline{r}}\rangle(t)$ some time later. The expectation value is a point in space with a trajectory. The wave function cannot be described by a single point because it has an extension in terms of the root mean square deviation, and this extension is $\Delta\hat{\underline{r}}(t)$ at time t. Because the equation for the time evolution of the expectation value has the same form as Newton's second law, one might ask whether the center of the wave packet obeys the laws of classical mechanics, but this is not quite the case. The right side of Ehrenfest's theorem (2.11) is - $\langle\nabla V(\hat{\underline{r}})\rangle$ and not -$[\nabla V(\underline{r})]_{\underline{r}=\langle\hat{\underline{r}}\rangle}$ where the gradient of the classical potential is evaluated at the position of the center of the wave packet. Thus, the center of mass behaves like a classical particle if - $\langle\nabla V(\hat{\underline{r}})\rangle$ = -$[\nabla V(\underline{r})]_{\underline{r}=\langle\hat{\underline{r}}\rangle}$ but otherwise it does not obey the laws of classical physics.

The question is therefore whether $\langle\nabla V(\hat{\underline{r}})\rangle \stackrel{?}{=} \nabla V(\underline{r})]_{\underline{r}=\langle\hat{\underline{r}}\rangle}$ holds. The potential can be expanded into a power series, and the gradient is just the derivative such that

$$V(x) = \sum_n V_n\, x^n \qquad\qquad \Rightarrow \qquad\qquad V'(x) = \sum_n n\, V_n\, x^{n-1}$$

holds in one dimension though the argument can be generalized to three dimensions. Inserting gives

$$\langle V'(\hat{x})\rangle = \sum_n n\, V_n\, \langle x^{n-1}\rangle \qquad\qquad [V'(x)]_{x=\langle x\rangle} = \sum_n n\, V_n\, \langle x\rangle^{n-1}$$

and shows that the two expressions are equal if $\langle x^{n-1}\rangle = \langle x\rangle^{n-1}$ for all n. This is the case for $n \leq 2$ as in the example $n = 2$ with $\langle x^1\rangle = \langle x\rangle^1$ but is not generally the case for $n > 2$ as in the example $n = 3$ with $\langle x^2\rangle \neq \langle x\rangle^2$. This means that the point associated with the center of the wave packet does in general not follow the rules of classical mechanics.

However, for $n \leq 2$ the point associated with the center of the wave packet follows the rules of classical mechanics. This includes several important types of potentials. The potential is constant for $n = 0$ and this corresponds to the free particle. In the case $n = 1$ the potential is linear $V(x) \sim x$ and the force field is uniform such as an external electric field. For $n = 2$ the potential is quadratic $V(x) \sim x^2$ as for the harmonic oscillator.

There is also the quasi-classical regime where the difference between $\langle\nabla V(\hat{\underline{r}})\rangle$ and $[\nabla V(\underline{r})]_{\underline{r}=\langle\hat{\underline{r}}\rangle}$ can be neglected. This is the case if the wave packet is sufficiently localized compared to other relevant length scales or, mathematically, if $|\psi(\underline{r},t)|^2 \neq 0$ in $\Delta\hat{\underline{r}}$ but $|\psi(\underline{r},t)|^2 \approx 0$ elsewhere. This is in particular the case when changes in the potential $V(\hat{\underline{r}})$ occur over length scales much larger than $\Delta\hat{\underline{r}}$ such that $\langle\nabla V(\hat{\underline{r}})\rangle \approx [\nabla V(\underline{r})]_{\underline{r}=\langle\hat{\underline{r}}\rangle}$. Thus,

$$\begin{aligned}\langle\nabla V(\hat{\underline{r}})\rangle &= \int d\underline{r}\,\psi^*(\underline{r},t)\big[\nabla V(\underline{r})\big]\psi(\underline{r},t) \approx \int d\underline{r}\,\psi^*(\underline{r},t)\big[\nabla V(\underline{r})\big]_{\underline{r}=\langle\hat{\underline{r}}\rangle}\psi(\underline{r},t)\\ &= \big[\nabla V(\underline{r})\big]_{\underline{r}=\langle\hat{\underline{r}}\rangle}\int d\underline{r}\,|\psi(\underline{r},t)|^2 = \big[\nabla V(\underline{r})\big]_{\underline{r}=\langle\hat{\underline{r}}\rangle}\end{aligned}$$

because $\psi(\underline{r},t)$ is normalized. This shows that the expectation value of the gradient of the potential is approximately equal to the gradient of the potential at the center of the wave packet in the quasi-classical

regime. In other word, the center of the wave packet behaves pretty much like a classical particle in this approximation.

The constraints $|\psi(\underline{r},t)|^2 \neq 0$ in $\Delta\hat{\underline{r}}$ and $|\psi(\underline{r},t)|^2 \approx 0$ elsewhere are met by most macroscopic objects because one can take the wave length $\lambda_{\text{de Broglie}}$ as the size of the wave packet $\Delta\hat{x}$ and this is much smaller than the distance over which the potentials vary. The quasi-classical regime is therefore very relevant.

2.9 The Time-Energy Uncertainty Principle

There is an uncertainty principle for any two observables in quantum mechanics, but time is not an observable in quantum mechanics because it is a parameter. This already suggests that the time-energy uncertainty principle is different from the other cases but it is not unrelated.

A system where $\hat{H}$ is time independent is called a conservative system. An observable $\hat{A}$ that is independent of time satisfies $\partial\hat{A}/\partial t = 0$. Using mean square deviation gives

$$\Delta\hat{H} = \sqrt{\langle\hat{H}^2\rangle - \langle\hat{H}\rangle^2} = \sqrt{\langle\hat{\sigma}_H^2\rangle} \qquad \sqrt{\langle\hat{\sigma}_H^2\rangle} = \hat{H} - \langle\hat{H}\rangle$$
$$\Delta\hat{A} = \sqrt{\langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2} = \sqrt{\langle\hat{\sigma}_A^2\rangle} \qquad \sqrt{\langle\hat{\sigma}_A^2\rangle} = \hat{A} - \langle\hat{A}\rangle$$

and inserting Ehrenfest's result (2.11) into the uncertainty principle leads to

$$\Delta\hat{H}\Delta\hat{A} \geq \frac{1}{2}\left|i\,\hbar\,\frac{d\,\langle\hat{A}\rangle}{dt}\right| = \frac{\hbar}{2}\left|\frac{d\,\langle\hat{A}\rangle}{dt}\right| \qquad \Rightarrow \qquad \left(\frac{\Delta\hat{A}}{\left|\frac{d\langle\hat{A}\rangle}{dt}\right|}\right)\Delta\hat{H} = \Delta T\,\Delta E \geq \frac{1}{2}\hbar$$

where the first factor has been renamed to ΔT and the second from $\Delta\hat{H}$ to ΔE. Thus, the time-energy uncertainty relation becomes $\Delta T\,\Delta E \geq \frac{1}{2}\hbar$.

The two uncertainty relations $\Delta T\,\Delta E \geq \frac{1}{2}\hbar$ and $\Delta\hat{x}\,\Delta\hat{p} \geq \frac{1}{2}\hbar$ look similar but they are fundamentally different. Position and momentum are observables represented by operators. Also energy as the Hamiltonian is an observable represented by an operator. However, time is not an observable but a parameter in quantum mechanics.

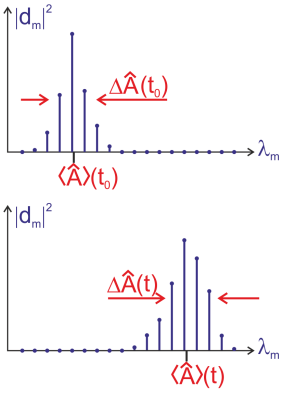
To understand the difference one uses

$$\hat{A}\,|v_m\rangle = \lambda_m\,|v_m\rangle \qquad |\psi\rangle = \sum_m d_m\,|v_m\rangle \qquad d_m = \langle v_m|\psi\rangle$$

such that $|\psi\rangle$ is expanded in the basis of eigenstates of $\hat{A}$. The time difference between the two situations

$$|\psi(t_0)\rangle = \sum_m d_m(t_0)\,|v_m\rangle \qquad |\psi(t)\rangle = \sum_m d_m(t)\,|v_m\rangle$$

is ΔT. The $\Delta\hat{A}$ in the numerator above is the width of the distributions as illustrated in the figure on the right side where it has been chosen rather small and of similar magnitude. The denominator has the time derivative $d\,\langle\hat{A}\rangle\,/dt$ of the mean $\langle\hat{A}\rangle$ which is a measure of the difference between the mean at the initial time t_0 compared to the final time t. This means that ΔT is large if it takes a long time for the distributions to change more than their width and it is very small if the distributions change rapidly compared to their width. In other words, ΔT measures how fast the state of the system changes.



Time evolution is governed in quantum mechanics by the Schrödinger equation with the Hamiltonian $\hat{H}$ whose eigenvalue equation is $\hat{H}\,|u_n\rangle = E_n\,|u_n\rangle$. The uncertainty equation $\Delta T\,\Delta E \geq \frac{1}{2}\hbar$ is saying that there is a relation between how fast a system can evolve in time and how it is written down in the energy basis. The system is in state

$$|\psi(t)\rangle = \sum_n c_n(t)\,|u_n\rangle \qquad c_n(t) = \langle u_n|\psi(t)\rangle$$

written in the energy basis and there are two different extremes:

(i) $\Delta E \ll 1$ means that the state of the system can be written with very few non-zero $c_n(t)$ and therefore $\Delta T \gg 1$ such that the system changes slowly.
(ii) $\Delta E \gg 1$ means that the state of the system can be written with many non-zero $c_n(t)$ and therefore $\Delta T \ll 1$ such that the time scale associated with changes in the system is very short and the system changes very fast.

Thus, if the system at time t_0 is an eigenstate of the Hamiltonian such that $|\psi(t_0)\rangle = |u_n\rangle$ then $\Delta E = 0$ because only one eigenvalue contributes. The change of the state from $|\psi(t_0)\rangle$ to $|\psi(t)\rangle$ by the Schrödinger equation is just the multiplication by a phase which is irrelevant for the physics. These states are called stationary. The expectation value for an observable $\hat{A}$ is the same at time t as at time t_0 and the time derivative $d\langle\hat{A}\rangle/dt$ is therefore zero such that ΔT goes to infinity according to the time-energy uncertainty relation.

2.10 Constants of Motion

Constants of motion are quantities that do not change as the system evolves in time. In classical mechanics, the angular momentum of a particle in a central potential, for example, is conserved. A consequence is that the particle moves in a plane. There are also constants of motion in quantum mechanics but the definition is more subtle.

An observable $\hat{A}$ is a *constant of motion* in quantum mechanics if it obeys the two conditions.

(1) $\hat{A}$ does not explicitly depend on time such that $\partial\hat{A}/\partial t = 0$.
(2) $\hat{A}$ commutes with the Hamiltonian such that $[\hat{A}, \hat{H}] = 0$.

The first condition is obvious and the second condition comes from the fact that time evolution in quantum mechanics is governed by the Schrödinger equation.

A first property of an observable that is a constant of motion is that its expectation value with respect to any state is time independent. From Ehrenfest's theorem (2.10) follows

$$\frac{d}{dt}\langle\hat{A}(t)\rangle = \frac{1}{i\hbar}\left\langle\left[\hat{A}(t), \hat{H}(t)\right]\right\rangle + \left\langle\frac{\partial}{\partial t}\hat{A}(t)\right\rangle = 0$$

and both terms are therefore zero because of the above conditions.

The second property of constants of motion is that they have a common set of eigenstates with $\hat{H}$ because of the second condition. The eigenvalue equations can be written as

$$\hat{H}\,|n,p\rangle = E_n\,|n,p\rangle \qquad\qquad \hat{A}\,|n,p\rangle = a_p\,|n,p\rangle$$

for $\hat{H}$ and $\hat{A}$, respectively. The eigenstates of the Hamiltonian are called stationary states because

$$|\psi(t_0)\rangle = |n,p\rangle \qquad \Rightarrow \qquad |\psi(t)\rangle = e^{-i\,E_n(t-t_0)/\hbar}\,|n,p\rangle$$

where $e^{-i\,E_n(t-t_0)/\hbar}$ is just a global phase factor that does not change the predictions of quantum mechanics. The initial state $|\psi(t_0)\rangle$ is also an eigenstate of $\hat{A}$ and therefore a measurement of $\hat{A}$ gives always a_p. The eigenvalues of a constant of motion such as a_p are called good quantum numbers.

The final property of constants of motion is that the probability of a measurement outcome is time independent because

$$|\psi(t_0)\rangle = \sum_{np} c_{np}(t_0)\,|n,p\rangle \qquad\qquad c_{np}(t_0) = \langle n,p|\psi(t_0)\rangle$$

$$|\psi(t)\rangle = \sum_{np} c_{np}(t)\,|n,p\rangle \qquad\qquad c_{np}(t) = e^{-i\,E_n(t-t_0)/\hbar}\,c_{np}(t_0)$$

and

$$\begin{aligned} P(a_p,t_0) &= |c_{np}(t_0)|^2 \\ P(a_p,t) &= |c_{np}(t)|^2 = \left|e^{-i\,E_n(t-t_0)/\hbar}\,c_{np}(t_0)\right|^2 \\ &= |c_{np}(t_0)|^2 \end{aligned}$$

show that $P(a_p, t) = P(a_p, t_0)$. Thus, if one measures a physical property that is a constant of motion then the probability of getting a particular eigenvalue is the same at any time.

To summarize, a constant of motion $\hat{A}$ has the properties:

(1) the expectation value $\langle \hat{A} \rangle$ is time independent,
(2) the eigenstates of $\hat{A}$ are time independent, and
(3) the probability of a measurement outcome of $\hat{A}$ is time independent.

Constants of motion in quantum mechanics are very useful because of the time independence of expectation values, of eigenstates and of measurement probabilities. As an example angular momentum is a constant of motion in central potentials.

2.11 Pure and Mixed States

Given are an operator $\hat{A}$ with two normalized eigenstates $|u_1\rangle$ and $|u_2\rangle$ corresponding to distinct eigenvalues λ_1 and λ_2 plus a superposition $|\psi\rangle = c_1 |u_1\rangle + c_2 |u_2\rangle$ of these eigenstates with $|c_1|^2 + |c_2|^2 = 1$. A measurement of $\hat{A}$ in state $|\psi\rangle$ gives either λ_1 with probability $P(\lambda_1) = |\langle u_1|\psi\rangle|^2 = |c_1|^2$ or λ_2 with probability $P(\lambda_2) = |\langle u_2|\psi\rangle|^2 = |c_2|^2$. The state changes immediately from $|\psi\rangle$ to $|u_1\rangle$ if λ_1 is the outcome of the measurement or to $|u_2\rangle$ if λ_2 is the outcome of the measurement.

In a framework based on a superposition state, the first step is to set up the system in the superposition state $|\psi\rangle = c_1 |u_1\rangle + c_2 |u_2\rangle$, and the second step is to select one of the states. Because $|\psi\rangle$ is here the only state, the selection in this case is deterministic and returns $|\psi\rangle$. The third step is the measurement of $\hat{A}$ where one can get λ_1 with probability $|c_1|^2$ or λ_2 with probability $|c_2|^2$. This step is probabilistic.

In a different framework called the statistical mixture of states the first step is also the setup of the system but in this case there are n_1 copies of $|u_1\rangle$ and n_2 copies of $|u_2\rangle$. If $|c_1|^2$ and $|c_2|^2$ are defined as

$$|c_1|^2 = \frac{n_1}{n_1 + n_2} \qquad\qquad |c_2|^2 = \frac{n_2}{n_1 + n_2}$$

the selection of one of the states in the second step gives $|u_1\rangle$ with probability $|c_1|^2$ and $|u_2\rangle$ with probability $|c_2|^2$. This step is therefore probabilistic. The measurement as the last step is now deterministic because λ_1 is measured if $|u_1\rangle$ has been selected and λ_2 is measured if $|u_2\rangle$ has been selected.

Both frameworks have a probabilistic and a deterministic step, and the outcome is the same. However, in the superposition framework the deterministic step is the second step and the probabilistic step is the third step while in the statistical framework the probabilistic step is the second step and the deterministic step is the third step. There is a deep difference in what probability means. In the superposition framework the probabilistic nature comes from the intrinsic properties of quantum theory, but in the statistical framework the probabilistic nature comes from the fact that the knowledge about the system is only partial and also happens in classical physics.

Despite the fact that the two frameworks give the same answer, the difference is fundamental. A system of two observables $\hat{A}$ and $\hat{B}$ with eigenvalue equations $\hat{A}|u_j\rangle = \lambda_j |u_j\rangle$ and $\hat{B}|v_k\rangle = \mu_k |v_k\rangle$ shows the difference. In the superposition framework with $|\psi\rangle = c_1 |u_1\rangle + c_2 |u_2\rangle$ the probability $P(\mu_k)$ is

$$\begin{aligned} P(\mu_k) &= |\langle v_k|\psi\rangle|^2 = |c_1 \langle v_k|u_1\rangle + c_2 \langle v_k|u_2\rangle|^2 = \left(c_1^* \langle v_k|u_1\rangle^* + c_2^* \langle v_k|u_2\rangle^*\right)\left(c_1 \langle v_k|u_1\rangle + c_2 \langle v_k|u_2\rangle\right) \\ &= |c_1|^2 |\langle v_k|u_1\rangle|^2 + |c_2|^2 |\langle v_k|u_2\rangle|^2 + \mathrm{Re}\{c_1 c_2^* \langle v_k|u_1\rangle \langle v_k|u_2\rangle p^*\} \end{aligned}$$

while in the statistical framework with probabilities $P(|u_1\rangle) = |c_1|^2$ and $P(|u_2\rangle) = |c_2|^2$ the probability to measure μ_k in case of $|u_1\rangle$ is $P_{u_1}(\mu_k) = |\langle v_k|u_1\rangle|^2$ and in case of $|u_2\rangle$ is $P_{u_2}(\mu_k) = |\langle v_k|u_2\rangle|^2$ such that

$$P(\mu_k) = |c_1|^2 |\langle v_k|u_1\rangle|^2 + |c_2|^2 |\langle v_k|u_2\rangle|^2$$

is the total probability to measure μ_k. The extra term $\mathrm{Re}\{c_1 c_2^* \langle v_k|u_1\rangle \langle v_k|u_2\rangle p^*\}$ of the superposition framework is not present in the statistical framework. This term is called the *interference term* and is essential for quantum mechanics. The double-slit experiment, for example, cannot be understood without this term. Thus, the superposition framework is the correct framework for describing quantum mechanics

although statistical mixtures may occasionally lead to the same predictions as quantum mechanics. The framework of statistical mixture of states is still important in quantum mechanics because one may encounter situations where one does not have enough information about the given system. However, the probabilistic step has nothing to do with the postulates of quantum mechanics.

If one knows everything one can know about the quantum system then probability is a purely quantum effect and the corresponding quantum state is called *pure state*. The state $|\psi\rangle = c_1 |u_1\rangle + c_2 |u_2\rangle$ used above, for example, is a pure state. If one has only partial knowledge about the system then probability enters in two places. There is the quantum mechanical probability of measurements, but also the lack of knowledge leads to probability to determine the state one is dealing with. The second type of probability happens also in classical physics. In this case the state is called a *mixed state*. A macroscopic system, for example, with in the order of 10^{23} particles is impossible to fully characterize, and one uses probability in the context of statistical mechanics. A system at thermodynamic equilibrium has a probability $P(E)$ to be in a state of energy E proportional to $e^{-E/k_B T}$ where k_B is Boltzmann's constant. In a mixed state the system is in state $|\psi_1\rangle$ with probability p_1, in state $|\psi_2\rangle$ with probability p_2 and so on with $\sum_j p_j = 1$ because the system is certainly in one of these states. Each of the states $|\psi_j\rangle$ is a pure state with certain probabilities when measured.

2.12 Density Operators for Pure States

To introduce the density operator for pure states a basis $\{|u_j\rangle\}$ for the state space is assumed that is orthonormal $\langle u_j|u_k\rangle = \delta_{jk}$. A state vector $|\psi\rangle$ can be specified as $|\psi\rangle = \sum_j c_j |u_j\rangle$ as usual. The *density operator* is defined as $\hat{\rho} = |\psi\rangle\langle\psi|$ for pure states, and $\hat{\rho}$ is therefore the projection operator onto $|\psi\rangle$. Written in the basis $\{|u_j\rangle\}$ the density operator $\hat{\rho}$ is

$$\hat{\rho} = |\psi\rangle\langle\psi| \qquad\qquad \rho_{jk} = \langle u_j|\hat{\rho}|u_k\rangle = \langle u_j|\psi\rangle \langle\psi|u_k\rangle = c_j\, c_k^* \tag{2.12}$$

and this shows that $|\psi\rangle$ as the state and $\hat{\rho}$ as the corresponding density operator contain the same information. These two quantities are

$$|\psi\rangle = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix} \qquad\qquad \rho_{jk} = \begin{pmatrix} c_1 c_1^* & c_1 c_2^* & \dots \\ c_2 c_1^* & c_2 c_2^* & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

in the matrix formulation of quantum mechanics.

One can develop quantum mechanic completely using state vectors but one can alternatively develop it using density operators instead. For pure states the version of quantum mechanics using states is simpler but for mixed states the version using density operators is more convenient.

The density operator has useful properties. It is Hermitian because $\hat{\rho}^\dagger = (|\psi\rangle\langle\psi|)^\dagger = |\psi\rangle\langle\psi| = \hat{\rho}$, and it is idempotent because $\hat{\rho}^2 = |\psi\rangle \langle\psi|\psi\rangle \langle\psi| = |\psi\rangle\langle\psi| = \hat{\rho}$ as $|\psi\rangle$ is normalized. Since the density operator for a pure state is a projection operator these two properties are no surprise. However, the density operator for a mixed state is not idempotent such that this property allows to distinguish between pure and mixed states. A third property is that there is no global phase ambiguity because $e^{i\theta} |\psi\rangle\langle\psi|\, e^{-i\theta} = |\psi\rangle\langle\psi|$, and this is an advantage compared to quantum mechanics in terms of states in a Hilbert space.

Normalization for a state vector means $\langle\psi|\psi\rangle = 1$ such that

$$\langle\psi|\psi\rangle = \left(\sum_j c_j^* \langle u_j|\right)\left(\sum_k c_k |u_k\rangle\right) = \sum_{jk} c_j^* c_k \langle u_j|u_k\rangle = \sum_{jk} c_j^* c_k \delta_{jk} = \sum_j |c_j|^2 = 1$$

and this corresponds to

$$\rho_{jj} = |c_j|^2 \qquad\qquad 1 = \sum_j |c_j|^2 = \sum_j \rho_{jj} = \mathrm{tr}(\hat{\rho})$$

for the density operator. The two expression $\langle\psi|\psi\rangle = 1$ and $\mathrm{tr}(\hat{\rho}) = 1$ are equivalent.

A quantum system is fully specified by the expectation values of a complete set of commuting observables. The expectation value of an observable can be built in terms of a density operator. The expectation value $\langle\hat{A}\rangle$ is

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

in terms of states. It can be written as

$$\langle\hat{A}\rangle = \sum_{jk}\langle\psi\,|u_j\rangle\langle u_j|\,\hat{A}\,|u_k\rangle\langle u_k|\,\psi\rangle = \sum_{jk}\langle u_k|\psi\rangle\,\langle\psi|u_j\rangle\,\langle u_j|\hat{A}|u_k\rangle = \sum_k\sum_j\langle u_k|\hat{\rho}|u_j\rangle\,\langle u_j|\hat{A}|u_k\rangle$$
$$= \sum_k\langle u_k|\,\hat{\rho}\left(\sum_j|u_j\rangle\langle u_j|\right)\hat{A}\,|u_k\rangle = \sum_k\langle u_k|\,\hat{\rho}\,\mathbb{I}\,\hat{A}\,|u_k\rangle = \sum_k\langle u_k|\,\hat{\rho}\hat{A}\,|u_k\rangle = \sum_k(\hat{\rho}\hat{A})_{jj} = \mathrm{tr}\Big(\hat{\rho}\hat{A}\Big)$$

in terms of density operators. The two expressions $\langle\hat{A}\rangle = \langle\psi|\hat{A}|\psi\rangle$ and $\langle\hat{A}\rangle = \mathrm{tr}\Big(\hat{\rho}\hat{A}\Big)$ are equivalent.

The probability for a measurement outcome is $P(\lambda_n) = \langle\psi|\hat{P}_n|\psi\rangle$ where the projection operator $\hat{P}_n$ projects onto the subspace of dimension g_n spanned by the eigenstates of the eigenvalue λ_n. In terms of states it can be written as

$$\hat{A}\,|u_n^j\rangle = \lambda_n\,|u_n^j\rangle \qquad j = 1, ..., g_n \qquad P(\lambda_n) = \langle\psi|\hat{P}_n|\psi\rangle \qquad \hat{P}_n = \sum_{j=1}^{g_n}\big|u_n^j\big\rangle\big\langle u_n^j\big|$$

and in terms of density operators as

$$P(\lambda_n) = \mathrm{tr}\Big(\hat{\rho}\hat{P}_n\Big)$$

using the above result for expectation values.

Time evolution is governed by the Schrödinger equation. The derivative of a density operator is

$$\frac{d}{dt}\hat{\rho}(t) = \frac{d}{dt}\,|\psi(t)\rangle\langle\psi(t)| = \left(\frac{d}{dt}\,|\psi(t)\rangle\right)\langle\psi(t)| + |\psi(t)\rangle\left(\frac{d}{dt}\,\langle\psi(t)|\right)$$
$$= \frac{1}{i\hbar}\hat{H}(t)\,|\psi(t)\rangle\langle\psi(t)| - \frac{1}{i\hbar}\,|\psi(t)\rangle\langle\psi(t)|\,\hat{H}^\dagger(t) = \frac{1}{i\hbar}\Big(\hat{H}(t)\hat{\rho}(t) - \hat{\rho}(t)\hat{H}(t)\Big) = \frac{1}{i\hbar}\Big[\hat{H}(t), \hat{\rho}(t)\Big]$$

using the chain rule and the Schrödinger equation. The two equations

$$\frac{d}{dt}\,|\psi(t)\rangle = \frac{1}{i\hbar}\hat{H}(t)\,|\psi(t)\rangle \qquad\qquad \frac{d}{dt}\hat{\rho}(t) = \frac{1}{i\hbar}\Big[\hat{H}(t), \hat{\rho}(t)\Big] \tag{2.13}$$

for time evolution are equivalent.

2.13 Density Operators for Mixed States

For pure states where one has perfect knowledge about the state of the system, quantum mechanics can be described equivalently in terms of states in the state space or in terms of density operators. For mixed states where one has only incomplete information about the system as in quantum statistical mechanics the description in terms of density operators is much more convenient.

In a mixed state the system could be in state $|\psi_1\rangle$ or in state $|\psi_2\rangle$ and so on where each of these states is a pure state but it is not know which one is the one for the system. It is only known that the probability of $|\psi_k\rangle$ is p_k. The system is in one of these states and therefore $\sum_k p_k = 1$. These probabilities p_k encode the lack of knowledge about the system and have nothing to do with quantum mechanics. Macroscopic systems with in the order of 10^{23} particles in statistical mechanics are examples as discussed above.

The key reason for the fact that quantum mechanics in terms of density operators for mixed states is more convenient than using state vectors is that the density operator appears in a linear manner while state vectors appear in a quadratic manner. The expectation value, for example, contains the density operator only once in $\langle \hat{A} \rangle = \mathrm{tr}\left(\hat{\rho}\hat{A}\right)$ but contains the state vector twice in $\langle \hat{A} \rangle = \langle \psi | \hat{A} | \psi \rangle$.

A mixed state consisting of states $|\psi_k\rangle$ with probability p_k such that $\sum_k p_k = 1$ can be calculated by applying the usual rules of quantum mechanics to the pure states $|\psi_k\rangle$ and then averaging over the different states. Given $\hat{A}\,|u_j\rangle = \lambda_j\,|u_j\rangle$ and one of the states $|\psi_k\rangle$ in the statistical mixture, the probability to measure the eigenvalue λ_j in state $|\psi_k\rangle$ is $P_k(\lambda_j) = \mathrm{tr}\left(\hat{\rho}_k\hat{P}_j\right)$ where $\hat{\rho}_k = |\psi_k\rangle\langle\psi_k|$. Thus, the total probability to measure λ_j in the system is

$$P(\lambda_j) = \sum_k p_k\, P_k(\lambda_j) = \sum_k p_k\ \mathrm{tr}\left(\hat{\rho}_k\hat{P}_j\right) = \mathrm{tr}\left(\Big(\sum_k p_k\,\hat{\rho}_k\Big)\hat{P}_j\right) = \mathrm{tr}\Big(\hat{\rho}\hat{P}_j\Big)$$

where

$$\hat{\rho} = \sum_k p_k\,\hat{\rho}_k = \sum_k p_k\ |\psi_k\rangle\langle\psi_k| \tag{2.14}$$

defines the density operator for a mixed state. The density operator of a mixed state is therefore the average of the density operators of the individual pure states building the statistical mixture.

The expression for the probability of the outcome λ_j is the same as the one for the pure state. This definition (2.14) has the effect that all the relevant equations have the same form for pure states and for mixed states. Because all individual $\hat{\rho}_k$ are Hermitian also $\hat{\rho}$ is Hermitian.

The definition (2.14) shows that the density operator $\hat{\rho}$ for mixed states is linear in the density operators $\hat{\rho}_k$ for the pure states. Because all the relevant quantities in quantum mechanics written in terms of the density operators for pure states are linear, these quantities generalize to mixed states in a simple manner. Normalization, for example, means for a pure state $\mathrm{tr}(\hat{\rho}_k) = 1$ and for a mixed state

$$\mathrm{tr}(\hat{\rho}) = \mathrm{tr}\left(\sum_k p_k\,\hat{\rho}_k\right) = \sum_k pk\ \mathrm{tr}(\hat{\rho}_k) = \sum_k p_k = 1$$

and therefore also $\mathrm{tr}(\hat{\rho}) = 1$.

The expectation value for a pure state is $\langle \hat{A} \rangle = \mathrm{tr}\left(\hat{\rho}_k\hat{A}\right)$. The operator $\hat{A}$ can be written as

$$\hat{A}\,|u_j\rangle = \lambda_j\,|u_j\rangle \qquad\qquad \hat{P}_j = |u_j\rangle\langle u_j| \qquad\qquad \hat{A} = \sum_j \lambda_j\hat{P}_j$$

because

$$\hat{A}\,|\psi\rangle = \hat{A}\sum_k c_k\,|u_k\rangle = \sum_k c_k\hat{A}\,|u_k\rangle = \sum_k c_k\ |u_k\rangle\,\lambda_k$$

$$\sum_j \lambda_j\hat{P}_j\,|\psi\rangle = \sum_j \lambda_j\,|u_j\rangle\langle u_j|\sum_k c_k\,|u_k\rangle = \sum_{jk} c_k\,\lambda_j\,|u_j\rangle\,\langle u_j|u_k\rangle = \sum_{jk} c_k\,\lambda_j\,|u_j\rangle\,\delta_{jk} = \sum_k c_k\ |u_k\rangle\,\lambda_k$$

and the expectation value for a mixed state is

$$\langle \hat{A} \rangle = \sum_j \lambda_j P(\lambda_j) = \sum_j \lambda_j\,\mathrm{tr}\Big(\hat{\rho}\hat{P}_j\Big) = \mathrm{tr}\left(\hat{\rho}\left(\sum_j \lambda_j\hat{P}_j\right)\right) = \mathrm{tr}\Big(\hat{\rho}\hat{A}\Big)$$

using this result.

For the time evolution of a mixed state it is assumed that the Hamiltonian $\hat{H}(t)$ is known. In this case

$$\frac{d}{dt}\hat{\rho}(t) = \frac{d}{dt}\left(\sum_k p_k\hat{\rho}_k(t)\right) = \sum_k p_k\frac{d}{dt}\hat{\rho}_k(t) = \sum_k p_k\left(\frac{1}{i\hbar}\left[\hat{H}(t),\hat{\rho}_k(t)\right]\right) = \frac{1}{i\hbar}\left[\hat{H}(t),\sum_k p_k\hat{\rho}_k(t)\right]$$

and this gives the time evolution

$$\frac{d}{dt}\hat{\rho}(t) = \frac{1}{i\hbar}\left[\hat{H}(t), \hat{\rho}(t)\right] \tag{2.15}$$

for the density operator of the mixed state.

With the density operators for pure states on the left side and for mixed states on the right side

$$\hat{\rho}_k = |\psi_k\rangle\langle\psi_k| \qquad\qquad \hat{\rho} = \sum_k p_k \, |\psi_k\rangle\langle\psi_k|$$

all relevant formula of quantum mechanics have the same form for both pure and mixed states. This is convenient but sometimes it is not clear whether the given density operator is for a pure or a mixed state. For a pure state the density operator is idempotent such that $\mathrm{tr}\left(\hat{\rho}_k^2\right) = \mathrm{tr}\left(\hat{\rho}_k\right) = 1$ because it is a projection operator. For the density operator of a mixed state this is not the case as

$$\begin{aligned}\hat{\rho}^2 &= \left(\sum_j p_j\hat{\rho}_j\right)\left(\sum_k p_k\hat{\rho}_k\right) = \left(\sum_j p_j \, |\psi_j\rangle\langle\psi_j|\right)\left(\sum_k p_k \, |\psi_k\rangle\langle\psi_k|\right) = \sum_{jk} p_j p_k \, |\psi_j\rangle \, \langle\psi_j|\psi_k\rangle \, \langle\psi_k| \\ &= \sum_{jk} p_j p_k \, |\psi_j\rangle \, \delta_{jk} \, \langle\psi_k| = \sum_j p_j^2 \, |\psi_j\rangle\langle\psi_j| = \sum_k p_j^2\hat{\rho}_k \neq \sum_k p_j\hat{\rho}_k = \hat{\rho}\end{aligned}$$

shows assuming that the $|\psi_j\rangle$ are orthonormal. The density operator for a mixed state is not a projection operator. Thus, this allows to distinguish between the density operator of a pure state and the density operator of a mixed state. As long as there are at least two values $p_k \neq 0$ as needed for a mixed state the trace satisfies

$$\mathrm{tr}\big(\hat{\rho}^2\big) = \mathrm{tr}\left(\sum_k p_k^2\hat{\rho}_k\right) = \sum_k p_k^2 \, \mathrm{tr}(\hat{\rho}_k) = \sum_k p_k^2 < 1$$

because $0 \leq p_k < 1$ follows from $\sum_k p_k = 1$ and $p_k \geq 0$ as required for probabilities.