Abstract. This is a sketch of an overview of the early development of quantum mechanics in roughly the first 30 years of the 20th century, clarifying some common points of confusion, and emphasizing things in it that a mathematician can care about and work on. They grew from my own efforts in sorting out this confusing subject while I was trying to learn it as a mathematician uninitiated in physics, using all sorts of vague texts that assumed all sorts of different backgrounds and conventions. A lot is still missing or imprecise and will be added over time.

Contents

1. Stuff before we begin 2
   1.1. Nobel prizes 2
   1.2. Numeric value of physical constants 2
2. Math Preliminaries 2
   2.1. Vector calculus 2
   2.2. Fourier Transforms 3
3. Classical Physics 3
   3.1. Electromagnetism 4
   3.2. Classical physics of particles 4
   3.3. Classical physics of waves 5
4. Origins of Quantum Physics 8
   4.1. Blackbody radiation 8
   4.2. Photoelectric effect 8
   4.3. Bohr atom and the Rydberg formula 9
   4.4. de Broglie’s matter waves 10
5. Establishment of Quantum Physics 12
   5.1. Born rule 12
   5.2. Gaussian Wave Packet 13
   5.3. Uncertainty Principle 13
   5.4. Matrix mechanics 14
   5.5. Schrödinger equation 15
   5.6. Hydrogen atom 17
6. Copenhagen interpretation 19
   6.1. Ehrenfest Theorem 21
   6.2. Uncertainty principle, algebraic version 22
   6.3. Pauli exclusion principle 23
7. Stability of Matter 23
8. Appendix: Operator Theory 23
9. References 29
1. **Stuff before we begin**

1.1. **Nobel prizes.** Some of the people that got physics Nobel prizes for work in quantum physics done in the period we’re talking about:

- 1911: Wilhelm Wien (German), *for his discoveries regarding the laws governing the radiation of heat.*
- 1918: Max Planck (German), *in recognition of the services he rendered to the advancement of Physics by his discovery of energy quanta.*
- 1919: Johannes Stark (German), *for his discovery of the Doppler effect in canal rays and the splitting of spectral lines in electric fields.*
- 1921: Albert Einstein (German), *for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect.*
- 1922: Niels Bohr (Danish), *for his services in the investigation of the structure of atoms and of the radiation emanating from them.*
- 1927: Arthur Compton (American), *for his discovery of the effect named after him.*
- 1929: Louis de Broglie (French), *for his discovery of the wave nature of electrons.*
- 1932: Werner Heisenberg (German), *for the creation of quantum mechanics, the application of which has, inter alia, led to the discovery of the allotropic forms of Hydrogen.*
- 1933: Erwin Schrödinger (Austrian) and Paul Dirac (English), *for the discovery of new productive forms of atomic theory.*
- 1945: Wolfgang Pauli (Austrian-born Swiss), *for the discovery of the Exclusion Principle, also called the Pauli principle.*
- 1954: Max Born (German), *for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wavefunction.*

1.2. **Numeric value of physical constants.** We use Gaussian units, which means everything is expressed in terms of mass, length and time. One big difference to the SI units is that in the latter there is a separate concept for the dimension of electric charge (Coulomb), while in the former charge can be given in $\text{mass} \cdot \text{length}^{-1}$, as the dimensionless quantity $1/4\pi$ in Gaussian units. As a rule of thumb, SI units are preferred among experimentalists and Gaussian units among theorists.

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of light</td>
<td>$c = 2.998 \times 10^8 \text{ m/s}$</td>
<td>Speed</td>
</tr>
<tr>
<td>Electron’s mass</td>
<td>$m = 9.109 \times 10^{-31} \text{ kg}$</td>
<td>Mass</td>
</tr>
<tr>
<td>Elementary charge</td>
<td>$e = 1.519 \times 10^{-14} \text{ kg}^{1/2} \text{ m}^{3/2} / \text{s}$</td>
<td>Charge</td>
</tr>
<tr>
<td>Planck constant</td>
<td>$h = 6.626 \times 10^{-34} \text{ kg m}^2 / \text{s}$</td>
<td>Action</td>
</tr>
<tr>
<td>Reduced Planck constant</td>
<td>$\hbar = 1.055 \times 10^{-34} \text{ kg m}^2 / \text{s}$</td>
<td>Action</td>
</tr>
<tr>
<td>Bohr radius</td>
<td>$a_0 = 5.292 \times 10^{-11} \text{ m}$</td>
<td>Length</td>
</tr>
<tr>
<td>Rydberg constant</td>
<td>$R_H = 1.097 \times 10^4 \text{ m}^{-1}$</td>
<td>Inverse length</td>
</tr>
<tr>
<td>Rydberg</td>
<td>$R_{yd} = 2.180 \times 10^{-18} \text{ kg m}^2 / \text{s}^2$</td>
<td>Energy</td>
</tr>
</tbody>
</table>

2. **Math Preliminaries**

- **Boldface symbols like $\mathbf{x}$** represent vectors in $\mathbb{R}^3$.
- The norm of a vector is denoted with normal print: $x = \|x\|$.
- The unit vector in the direction of $\mathbf{x}$ is denoted by a hat: $\hat{x} = \mathbf{x}/x$, in case $x \neq 0$.
- Following practices from physics, dots over a symbol, like $\dot{x}$ or $\ddot{x}$, will denote time derivative, and inner products $\langle x, y \rangle$ in Hilbert spaces will be linear on the second argument, antilinear on the first.
- A silly little thing that helps a lot with preventing sign errors: the inverse of $i$ is $-i$.

2.1. **Vector calculus.**

- The 3 differential operators: Scalar $\nabla \rightarrow$ Vector $\nabla \times$ Vector $\nabla \cdot$ Vector $\rightarrow$ Scalar
- Gradient: $\nabla f = (\partial_x f, \partial_y f, \partial_z f)$
- Curl: $\nabla \times \mathbf{V} = (\partial_z V_y - \partial_y V_z, \partial_x V_z - \partial_z V_x, \partial_y V_x - \partial_x V_y)$
- Divergence: $\nabla \cdot \mathbf{V} = \partial_x V_x + \partial_y V_y + \partial_z V_z$
2.2. Fourier Transforms. The Fourier transform $\mathcal{F}_a \psi$ with parameter $a > 0$ of an $L^1$ function $\psi : \mathbb{R}^N \to \mathbb{C}$ will be defined as the continuous function

$$\label{fformula} (\mathcal{F}_a \psi)(\mathbf{k}) = \frac{1}{(2\pi a)^{N/2}} \int_{\mathbb{R}^N} \psi(\mathbf{x}) e^{-\frac{i}{a} \mathbf{k} \cdot \mathbf{x}} d^N \mathbf{x} \quad , \quad \mathbf{k} \in \mathbb{R}^N$$

Some properties of this:

- The inverse $\mathcal{F}_a^{-1}$ is defined on $L^1$ functions by

$$\label{inversionformula} (\mathcal{F}_a^{-1} \phi)(\mathbf{x}) = \frac{1}{(2\pi a)^{N/2}} \int_{\mathbb{R}^N} \phi(\mathbf{k}) e^{\frac{i}{a} \mathbf{k} \cdot \mathbf{x}} d^N \mathbf{k} = (\mathcal{F}_a \phi)(-\mathbf{x})$$

This is proved by approximating $\phi$ with well-behaved functions so that Fubini can be used, but it’s not exactly trivial.

- $\mathcal{F}_a$ extends to an isometry of $L^2(\mathbb{R}^N; \mathbb{C})$:

$$\int |\psi|^2 = \int |\mathcal{F}_a \psi|^2 \quad , \quad \psi \in L^2$$

This is called the Plancherel Formula and follows easily from knowing the inverse formula. So $\mathcal{F}$ is defined on all of $L^2$, although formula (1) only defines it on $L^1 \cap L^2$ functions. There’s no general formula in $L^2$ that doesn’t directly make use of distributions.

- $\mathcal{F}_a$ transforms rescaling of the independent variable into the inverse scaling of both dependent and independent variables: if $\psi_\lambda(x) = \psi(\lambda x)$ then

$$\frac{(\mathcal{F}_a \psi_\lambda)(k)}{\lambda} = \frac{1}{\lambda} \left( \frac{k}{\lambda} \right) \frac{(\mathcal{F}_a \psi)(k)}{\lambda} \quad , \quad \lambda > 0$$

This is a simple change of variable in the integral.

- $\mathcal{F}_a$ transforms derivatives into multiplication by polynomials. For $N = 1$, this means:

$$\mathcal{F}_a(\psi')(k) = \frac{i}{a} k (\mathcal{F}_a \psi)(k)$$

This is proved by integration by parts.

- $\mathcal{F}_a$ and $\mathcal{F}_a^{-1}$ transform translation into multiplication by exponential:

$$\mathcal{F}_a(\psi(\cdot - \mathbf{x}_0))(\mathbf{k}) = e^{-\frac{i}{a} \mathbf{k} \cdot \mathbf{x}_0} (\mathcal{F}_a \psi)(\mathbf{k}) \quad , \quad \mathcal{F}_a^{-1}(\phi(\cdot - \mathbf{k}_0))(\mathbf{x}) = e^{\frac{i}{a} \mathbf{a}_0 \cdot \mathbf{x}} (\mathcal{F}_a^{-1} \phi)(\mathbf{x})$$

This is a direct computation.

We’ll refer to $\mathcal{F}_1$ as simply $\mathcal{F}$, the standard Fourier transform. From the definition it’s clear that

$$\frac{(\mathcal{F}_a \psi)(\mathbf{p})}{\sqrt{a^N}} = (\mathcal{F}_1 \psi)(\frac{\mathbf{p}}{a})$$

3. Classical Physics

This section is intended to show examples of what particles and waves are in classical physics. The things that we calculate here will all have their use later when we need to talk about their counterparts in quantum physics.
3.1. **Electromagnetism.** Electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ are vector fields with (Gaussian) units of

\[ \text{[mass]}^{1/2} \cdot \text{[length]}^{-1/2} \cdot \text{[time]}^{-1} \]

generated from charge distributions and currents. They affect charged particles by the *Lorentz force*: a particle of charge $q$ and velocity $\mathbf{\dot{x}}$ experiences a force equal to

\[ \mathbf{F} = q \left( \mathbf{E} + \frac{\mathbf{\dot{x}}}{c} \mathbf{B} \right) \]

Let $\rho$ and $\mathbf{J}$ be a scalar and a vector field representing charge density and current density in space. The fields they generate satisfy the following laws:

- **Gauss’ Laws.** There are no magnetic monopoles (no flux of $\mathbf{B}$ through any closed surface), but flux of electric field through the boundary of a region is related to amount of charge inside:

\[
\oint_{\partial \Omega} \mathbf{B} \cdot d\mathbf{S} = 0 , \quad \frac{1}{4\pi} \oint_{\partial \Omega} \mathbf{E} \cdot d\mathbf{S} = \iiint_{\Omega} \rho \ dV
\]

- **Faraday’s Law of Induction.** Electric current around a loop is generated by a varying magnetic field inside:

\[
\oint_{\partial S} \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{c} \frac{d}{dt} \iint_{S} \mathbf{B} \ dS
\]

- **Ampere’s Law with Maxwell’s Correction.** Magnetic current around a loop is generated by a current inside and also by a varying electric field inside:

\[
\oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = 4\pi \frac{d}{dt} \iint_{S} \mathbf{J} \ dS + \frac{1}{c} \frac{d}{dt} \iint_{S} \mathbf{E} \ dS
\]

Due to the double and triple integral theorems of vector calculus applied to any arbitrary region of space or the plane, these can be written as the system of *Maxwell equations*

\[
\begin{cases}
\nabla \cdot \mathbf{B} = 0 \\
\nabla \cdot \mathbf{E} = 4\pi \rho \\
\n\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial}{\partial t} \mathbf{B} = 0 \\
\n\nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{E} = \frac{4\pi}{c} \mathbf{J}
\end{cases}
\]

3.2. **Classical physics of particles.** The Lorentz force gives us something to play with that has to do with how classical physics treats its particles. Suppose a big charge $Q$ sits at the origin, and another small charge $q$ of mass $m$ (a *test particle*) is under its influence. The electric and magnetic fields generated by the central charge at a point $\mathbf{r}$ are

\[ \mathbf{E} = \mathbf{E}(\mathbf{r}) = \frac{Q}{r^3} \mathbf{r} , \quad \mathbf{B} = 0 \]

This is obtained by solving the Maxwell equations for zero current density (the central charge is not going anywhere) and for a charge density of the type of a $\delta$ measure (the central charge is a point). If $\mathbf{r}$ denotes the position of the test particle (so that in everything below $\mathbf{r} = \mathbf{r}(t)$), the Lorentz force on it is

\[ \mathbf{F} = \frac{Qq}{r^3} \mathbf{r} \]

(also called the Coulomb force). This is repulsive if and only if the signs of the charges are the same, and the $r^3$ downstairs together with the $\mathbf{r}$ in front makes sure the magnitude of this force is inversely proportional to $r^2$, which makes it similar to Newtonian gravity when it’s attractive. It is also the gradient of a potential, which means there exists $V(\mathbf{r})$ such that $\mathbf{F} = -\nabla V$. Indeed, for

\[ V(\mathbf{r}) = \frac{Qq}{r} \]

we have

\[
\nabla V = Qq \nabla \frac{1}{\sqrt{r^2}} = Qq \left( -\frac{1}{2} (r^2)^{-3/2} \right) \nabla r^2 = -\left( \frac{Qq}{2r^3} \right) 2\mathbf{r} = -\frac{Qq}{r^3} \mathbf{r}
\]

With forces given by potentials like this, we can define the *energy* as the sum of kinetic and potential energy:

\[ E = \frac{m\mathbf{r}^2}{2} + V(\mathbf{r}) \]
(not to be confused with electric field \( \mathbf{E} \)), and, given Newton’s second law \( \mathbf{F} = m\ddot{\mathbf{r}} \), we see the energy is a conserved quantity of the motion:

\[
\frac{dE}{dt} = \frac{m}{2} 2\dot{r} \cdot \ddot{r} + \nabla V(\mathbf{r}) \cdot \dot{\mathbf{r}} = \dot{\mathbf{r}} \cdot \mathbf{F} - \mathbf{F} \cdot \dot{\mathbf{r}} = 0
\]

The momentum of the particle is defined as

\[ p = m \dot{\mathbf{r}} \]

and is clearly not conserved (the velocity is usually changing magnitude and direction all the time). A more suitable vector for central forces like this, which is conserved in time, is the angular momentum

\[ \mathbf{L} = m \mathbf{r} \times \dot{\mathbf{r}} \]

Indeed,

\[
\frac{d\mathbf{L}}{dt} = m \dot{\mathbf{r}} \times \dot{\mathbf{r}} + m \mathbf{r} \times \ddot{\mathbf{r}} = 0
\]

since \( \ddot{\mathbf{r}} \) is proportional to \( \mathbf{F} \) which is a multiple of \( \mathbf{r} \).

Assume the Coulomb force is attractive now (\( Qq < 0 \)). The simplest type of orbit of the test particle around the center is a circle. Assume it is a circle of constant radius \( r = R \). Differentiating once,

\[ r^2 = \text{const} \Rightarrow \mathbf{r} \cdot \dot{\mathbf{r}} = 0 \]

This tells us we can compute \( \mathbf{L} = |\mathbf{L}| \) by simply multiplying the absolute values of the two vectors in its definition:

\[ \mathbf{L} = m r \dot{r} \]

Differentiating another time we can find \( \dot{r} \) in terms of \( r \):

\[ \mathbf{r} \cdot \ddot{\mathbf{r}} = 0 \Rightarrow \dot{r}^2 + \mathbf{r} \cdot \ddot{\mathbf{r}} = 0 \Rightarrow \dot{r}^2 = -\frac{Qq}{mr^3} \mathbf{r} \cdot \mathbf{r} = \left| \frac{Qq}{m} \right| \]

This allows us to find the energy and the angular momentum in terms of \( R \):

\[
\begin{align*}
E &= \frac{m\dot{r}^2}{2} + \frac{Qq}{r} = \left| \frac{Qq}{m} \right| r - \left| \frac{Qq}{m} \right| R \\
L^2 &= m^2 r^2 \dot{r}^2 = m^2 r^2 \left| \frac{Qq}{m} \right| r = |Qq| mr
\end{align*}
\]

These will be used later. For now, with a positively charged particle at the center and a negatively charged electron moving in space, notice how it’s always possible to place the electron in a configuration in order to obtain any real value we want for its energy. Indeed, first rewrite the energy in the form

\[ E = \frac{\dot{p}^2}{2m} + \frac{Qq}{r} \]

To obtain very negative values, let the electron orbit closer to the center (look at the \( E \) formula in (2) above); to obtain positive values, give it very high momentum away from the center (it will not be bound in an orbit). The situation in quantum physics will be drastically different, basically because, in there, position and momentum cannot be freely and independently chosen like this.

3.3. Classical physics of waves. The Maxwell equations also give us something to play with that has to do with how classical physics treats its waves. Here they are again:

\[
\begin{align*}
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \cdot \mathbf{E} &= 4\pi \rho \\
\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} &= 0 \\
\nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} &= \frac{4\pi}{c} \mathbf{J}
\end{align*}
\]

The first equation implies \( \mathbf{B} = \nabla \times \mathbf{A} \) for some vector potential \( \mathbf{A} \). Then the third equation is

\[
\nabla \times \mathbf{E} + \frac{1}{c} \nabla \times \frac{\partial \mathbf{A}}{\partial t} = 0 \Rightarrow \mathbf{E} = -\nabla \varphi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \text{ for some scalar potential } \varphi
\]
This already deals with the two homogeneous equations. Using the \textit{curl of a curl} formula, the fourth equation becomes

\[ \nabla(\nabla \cdot A) - \Delta A + \frac{1}{c} \nabla \partial_t \varphi + \frac{1}{c^2} \partial_t^2 A = \frac{4\pi}{c} J \]

It’s actually possible to choose \( A \) and \( \varphi \) such that

\[ \nabla \cdot A + \frac{1}{c} \partial_t \varphi = 0 \]

which transforms fourth equation into a wave equation for (each component of) \( A \):

\[ \frac{1}{c^2} \partial_t^2 A - \Delta A = \frac{4\pi}{c} J \]

This is called the \textit{Lorenz gauge condition}. Indeed, the defining property for \( A \), which is \( B = \nabla \times A \), is kept unchanged if we replace \( A \to A + \nabla f \) for any \( f \). Then the defining property for \( \varphi \), which is \( E = -\nabla \varphi - \frac{1}{c} \partial_t A \), will be kept unchanged if we also replace \( \varphi \to \varphi - \frac{1}{c} f \). Under this change the expression in (3) changes into (use \( \nabla \cdot \nabla f = \Delta f \))

\[ \nabla \cdot A + \Delta f + \frac{1}{c} \partial_t \varphi - \frac{1}{c^2} \partial_t^2 f \]

which can be made equal to 0 by solving a wave equation for \( f \):

\[ \frac{1}{c^2} \partial_t^2 - \Delta = \nabla \cdot A + \frac{1}{c} \partial_t \varphi \]

So the Lorentz gauge makes the fourth equation a wave equation for \( A \). Under this gauge, the second equation is also a wave equation for \( \varphi \):

\[ -\Delta \varphi - \frac{1}{c} \nabla \cdot \partial_t A = 4\pi \rho \iff -\Delta \varphi - \frac{1}{c} \partial_t \left( -\frac{1}{c} \partial_t \varphi \right) = 4\pi \rho \iff \frac{1}{c^2} \partial_t^2 \varphi - \Delta \varphi = 4\pi \rho \]

The solutions will now depend on \( \rho \) and \( J \), and are complicated to write down with this generality. Let’s solve them in electromagnetic vacuum (\( \rho = J = 0 \)). For any vector \( k \neq 0 \), solutions can be obtained as follows:

\[
\begin{aligned}
\varphi(x, t) &= \varphi_0 e^{i(k \cdot x - ct)} \\
A(x, t) &= A_0 e^{i(k \cdot x - ct)}
\end{aligned}
\]

Indeed, this is because

\[
\partial_t e^{i(k \cdot x - ct)} = -ik e^{i(k \cdot x - ct)} \Rightarrow \partial_t^2 e^{i(k \cdot x - ct)} = -k^2 e^{i(k \cdot x - ct)}
\]

\[
\nabla e^{i(k \cdot x - ct)} = ik e^{i(k \cdot x - ct)} \Rightarrow \Delta e^{i(k \cdot x - ct)} = -k^2 e^{i(k \cdot x - ct)}
\]

The fact that we must take the complex exponentials for \( \varphi \) and \( A \) to be in phase with each other comes from the gauge condition (3), which relates \( \varphi \) to \( A \) and also gives a relation between their amplitudes:

\[
0 = \nabla \cdot A + \frac{1}{c} \partial_t \varphi = (iA_0 \cdot k - i\varphi_0 k)e^{i(k \cdot x - ct)} \Rightarrow \varphi_0 = \frac{1}{k} A_0 \cdot k
\]

Now calculate the fields using the \textit{curl of a product} formula:

\[
B = \nabla \times A = i k \times A_0 e^{i(k \cdot x - ct)}
\]

\[
E = -\nabla \varphi - \frac{1}{c} \partial_t A = (-i\varphi_0 k + i k A_0) e^{i(k \cdot x - ct)}
\]

\[
= i \left( -(A_0 \cdot k) \frac{k}{k} + k A_0 \right) e^{i(k \cdot x - ct)}
\]

Then \( E \) and \( B \) are perpendicular to each other (\( B \) is perpendicular to both \( A_0 \) and \( k \), while \( E \) is in the plane spanned by them), and also perpendicular to the vector \( k \) (this is clear for \( B \) and follows for \( E \) by computing \( -(A_0 \cdot k) \frac{k}{k} + k A_0 \cdot k = 0 \)). More general solutions can be found by superimposing such solutions for different amplitudes \( A_0, \varphi_0 \) and wave-vectors \( k \). Due to Fourier analysis the general solution is of this form (that’s really why we considered complex exponentials to begin with), but unless we know boundary

\footnote{This is not a typo. Lorentz and Lorenz are two different people.}
conditions we can’t say if it’s a discrete (Fourier series) or continuous (Fourier transform) superposition. These complex exponentials represent waves; let’s talk about waves.

A wave is an oscillating and propagating thing (vector or scalar). For our purposes, we can call something a wave if its components are given by the formula

\[ A(\mathbf{x}, t) = A_0 e^{i(k \cdot \mathbf{x} - \omega(k)t)} \]

for some amplitude \( A_0 \in \mathbb{C} \), wave-vector \( \mathbf{k} \in \mathbb{R}^3 \) and angular frequency \( \omega \). If the wave is supposed to be real-valued we can always take real or imaginary parts. For now assume \( \omega \) is a multiply of \( k \):

\[ \omega(k) = vk \quad , \quad v \in \mathbb{R}^+ \]

Formula (4) is called a plane wave or also a wave packet. It represents something that propagates in the direction of \( \mathbf{k} \) with speed \( v \), since

\[ A(x_0, t_0 + t) = A_0 e^{i(k \cdot x - vkt - vkt_0)} = A(x_0 - vt \hat{k}, t_0) \]

which should be thought of as “what the wave will be at a certain position \( t \) units of time in the future is what it is in the present \( vt \) units of distance behind that position, in the direction of \( \mathbf{k} \).”

Therefore Maxwell showed that free electromagnetic radiation naturally propagates through vacuum, without a medium to carry it, shooting like light rays towards any particular direction. Light is simply electromagnetic fields freely propagating.

The mathematical periods of the oscillation for a fixed time (wavelength) and for a fixed position (wave period) are

\[
\lambda = \frac{2\pi}{k} \quad , \quad T = \frac{2\pi}{v} = \frac{2\pi}{\omega}
\]

Check this by noticing that replacing \( x \mapsto x + (2\pi/k)\hat{k} \) or \( t \mapsto t + 2\pi/\omega \) produces the same value for \( A \). The frequency of the wave is defined as

\[
\nu = \frac{1}{T} = \frac{\omega}{2\pi}
\]

A dissipating wave packet can be obtained by superimposing waves of different wavelengths (different values for \( k \)) that each propagate at a different speed (different values for \( v \)). This can be taken into account by letting now \( v = v(k) \):

\[ A(\mathbf{x}, t) = A_0 e^{i(k \cdot \mathbf{x} - v(k)t)} \]

The formula \( \omega(k) = v(k)k \) is called the dispersion relation of this plane wave. The case \( v(k) = \text{const} \) corresponds to free dispersion. So a general plane wave \( \psi(\mathbf{x}, t) \) can be written as a continuous superposition of wave packets, with amplitudes \( \phi(\mathbf{k}) \) that vary with the wave-vector \( \mathbf{k} \). Taking \( t = 0 \) for simplicity, this means:

\[
\psi(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int \phi(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}} d^3k
\]

Then this is simply the inverse Fourier transform of \( \psi \), and the amplitude \( \phi \) can be recovered by \( \mathcal{F} \):

\[
\phi(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int \psi(\mathbf{x}) e^{-i\mathbf{k} \cdot \mathbf{x}} d^3x
\]

Allowing time to vary, we can contemplate expressions such as

\[
\psi(\mathbf{x}, t) = \frac{1}{(2\pi)^{3/2}} \int \phi(\mathbf{k}) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega(k)t)} d^3k
\]

This will give rise to a wave that changes its form as it propagates, because each individual packet that forms it propagates at a different speed.
4. Origins of Quantum Physics

This section lists key experiments and theories developed in the first decades of the 20th century, known today as the old quantum theory. These ideas never attained the form of a self-consistent theory, but they helped create modern quantum theory\(^2\). The main concepts behind them are:

- **Quantization**: The value of some physical property of a system can only assume a discrete set of values.
- **Wave-particle duality**: Waves exhibit particle-like properties and vice-versa.

4.1. Blackbody radiation. A blackbody is an object that absorbs all radiation falling onto it. Radiation emitted from it can only come from itself. It’s an idealization imagined in order to study emission of electromagnetic radiation by solids when heated. They emit light of all frequencies, but with different intensities for each, depending on their temperature. Let \(u(\nu, T)\) be the intensity of light of frequency \(\nu\) emitted at temperature \(T\). This curve is usually sharply concentrated around a certain frequency, which determines the color that the object will appear to have when we look. This is how we know the temperature of stars and why hot metalwork glows red.

Stefan, Boltzmann and Wien, from 1884 to 1899, used empirical evidence and classical thermodynamics to obtain a formula for \(u(\nu, T)\) that turned out to be good only for high frequencies. Rayleigh (1900) and Jeans (1905) considered the same classical principles, but applied to the typical laboratory example of a blackbody (a solid cube with a small cavity inside, where the measurements of radiated heat are taken), and got a formula for \(u(\nu, T)\) that worked well only for low frequencies, predicting a divergent behavior at higher frequencies (the ultraviolet catastrophe).

The blackbody property for such a laboratory object with a cavity can be interpreted as stipulating that the only energy waves emitted by the object inside the cavity that do not end up being reabsorbed by it have to be standing waves, which mathematically means their frequencies can only be in a certain discrete set so that the wavelength divides the extent of the cavity, making the wave appear to not move, only oscillate in place. This makes it so that the wave is not really moving towards the object, so it’s not being absorbed by it. Based on this observation, Planck (1900) came up with his idea of quantization: he assumed that energy exchange between matter (the black body) and electromagnetic waves (radiation emitted by it) only happens in certain discrete multiples of the wave frequency \(\nu\). The constant of proportionality is today known as Planck’s constant \(h\):

\[
E = nh\nu, \quad n \in \mathbb{N}
\]

This idea and the formula for \(u(\nu, T)\) it yielded turned out to provide a good interpolation between the Wien and Rayleigh formulas and successfully explain blackbody radiation.

4.2. Photoelectric effect. Every metal has a threshold frequency \(\nu_0\) such that, if light of frequency \(\nu \geq \nu_0\) hits it, it emits electrons. The kinetic energy of the electrons depend on the frequency, not amplitude of the light. Higher amplitude only causes more electrons to be emitted.

The fact that electrons are emitted means the light is giving them enough energy to break from their nuclei. But if light is to be thought of as a classical wave, the observed effect contradicts what should be expected. The intensity and energy carried by the light should be directly related to its amplitude, not frequency. There would be no threshold frequency; electrons would always be emitted, just with a lower energy the lower the amplitude is.

Einstein (1905) followed on Planck’s quantization idea and proposed that light is composed of particles (latter named photons by the chemist Gilbert Lewis), each with energy

\[
E = h\nu
\]

When a photon hits an electron, if it has enough energy to scatter it, it’s absorbed and gives its energy to the electron. Otherwise, it just bounces off and keeps going. Light with higher intensity simply has more photons, but if the frequency is not high enough, none of them will have enough energy to break any electrons free. This simple idea explains every aspect of the photoelectric effect, but goes completely against the classical viewpoint that light is a wave, so it wasn’t immediately accepted by the physics community.

\(^2\)And whether or not the latter is self-consistent is a matter of much controversy.
The wave-like nature of light is not completely abandoned by Einstein. He postulates that there are light particles, but their energy is defined in terms of their frequency. That’s a property of a wave; what should the frequency of a particle be? Old quantum theory would never give a complete answer to this question.

The photoelectric effect is not to be confused with the Compton effect, which was discovered in 1923 and consists of electrons scattering photons, also demonstrating the particle-like nature of light.

4.3. Bohr atom and the Rydberg formula. This subsection is about Bohr’s model of the atom, which explained two separate aspects of their physics: spectral lines and stability of matter.

- **Spectral lines:** A gas in an excited state emits light, but not of all frequencies: only a discrete set of frequencies is observed, which depends on the initial excitation and on the chemical elements of the gas. Due to refraction, light of the different emitted frequencies travels at different angles and can be captured at different spots on an appropriately positioned screen, where it shows up as a series of lines of different colors (in case it’s visible light), that is, different frequencies (or wavelengths). The different values of the wavelengths obtained are called the spectral series of the particular element. By the end of the 19th century, tables existed showing the series of different wavelengths corresponding to many elements and their initial excitation states. Also some patterns had been deduced in each case and even used to predict new lines in some series. But no unifying formula or explanation was known for this phenomenon.

  In 1887 Rydberg empirically found a simple formula encompassing all of the different series. For Hydrogen specifically, it gives the wavelength $\lambda$ of a spectral line in terms of two positive integers $n_1, n_2$, with $n_2 \geq n_1 + 1$, and of an experimentally found constant $R_H$, the Rydberg constant:

  $$\frac{1}{\lambda} = R_H \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$$

  The parameter $n_1$ is related to the initial excitation state, and for a fixed $n_1$ the wavelengths of each spectral line are obtained by letting $n_2$ vary. But still there was nothing in physics to suggest where this phenomenon was coming from, let alone some theory able to produce this formula.

- **Stability of matter:** Rutherford discovered the atomic nucleus in 1911 and proposed that electrons orbited around it due to electric attraction. But if this were true, atoms would be sources of infinite energy, since the energy $(2)$ of an electron in circular orbit decreases to $-\infty$ as the radius decreases:

  $$E = -\frac{e^2}{2R}$$

  (this is for Hydrogen, where the electron has charge $-e$ and the nucleus $+e$). Particles colliding with the electron could rob some of its energy, relegating it to a closer orbit, and run away with it, and in principle this could keep happening forever. Or worse, a simple integration shows that an electron that constantly radiates electromagnetic energy due to its orbital motion would decay into the nucleus in a finite time (a fraction of a millisecond). In other words, matter would not be stable.

  In 1913 Bohr proposed a teleological postulate that solved this problem and explained the Rydberg formula. He stipulated that the radius of an electron orbiting a positively charged nucleus is quantized, that is, only a discrete set of values are allowed for it. When the atom receives or gives out energy, the electron instantaneously jumps to a higher or lower allowed orbit, respectively. The way that it gives out energy is by electromagnetic radiation, that is, by emitting photons. This means the possible values of energy emitted or absorbed are also discrete, and combining this with Einstein’s idea that energy is related to frequency in the case of light, Bohr’s theory could explain the discreteness of spectral emission lines. This was still a phenomenological theory; it didn’t say why the electrons should behave like this. It was proposed precisely to explain that.

  Quantitatively, what Bohr proposed was actually quantization of the angular momentum, from which quantization of the radius and of energy also follow. He stipulated that the electron’s angular momentum can only assume the values

  $$L = nh \quad , \quad n = 1, 2, 3, \ldots$$
Let \( m \) denote the mass of the electron. In (2) we found
\[
E = \frac{|Qq|}{2R} = -\frac{e^2}{2R}, \quad L^2 = |Qq|mR = me^2R
\]
With the Bohr quantization rule for \( L \), we have
\[
n^2\hbar^2 = me^2R \quad \implies \quad R = n^2\frac{\hbar^2}{me^2} = a_0n^2
\]
for the constants
\[
a_0 = \frac{\hbar^2}{me^2} \text{ (Bohr radius) }, \quad R_{yd} = \frac{me^4}{2\hbar^2} \text{ (Rydberg unit of energy) }
\]
The Bohr radius is the closest that an electron can get to the nucleus, so this model explains the stability of matter in a rather unsatisfactory way: it simply imposes that the electron should not be able to get closer because there are no allowed orbits there. The Rydberg unit of energy \( R_{yd} \), also called simply “Rydberg”, is not the Rydberg constant \( R_H \), but is closely related; in fact the Bohr model can also compute \( R_H \) in terms of fundamental constants. Let’s do that now.

An electron jumping from an orbit of radius \( a_0n_1^2 \) to another of radius \( a_0n_2^2 \) will emit a photon of energy
\[
E = R_{yd}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)
\]
We assume Einstein’s relation between energy and frequency, \( E = h\nu \). Since the wavelength \( \lambda \) is given as \( c/\nu \) (this is just the velocity formula \( c = \lambda/\nu \) together with \( \nu = 1/T \)), we get
\[
\frac{1}{\lambda} = \frac{E}{hc} = \frac{R_{yd}}{hc}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) = \frac{2\pi^2me^4}{\hbar^3c}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)
\]
so that
\[
R_H = \frac{2\pi^2me^4}{\hbar^3c}
\]
Check for yourself that this matches the experimental value!

4.4. de Broglie’s matter waves. Quantization of energy and photons are examples of the particle-like nature of waves. But around this time people were also contemplating (and verifying) that matter could behave like waves.

The double-slit experiment consists of a beam of light or matter passing through two thin slits on a surface and then hitting a screen and leaving some mark there. It was performed first with light rays in the early 1800’s by Young. A beam of classical particles would hit the screen only on locations aligned with the source and the slits, but a beam of classical waves would hit it in a broader area because it gets diffracted at the slits. It would also hit some locations with more intensity than others since the two diffracted waves formed at the slits interfere with each other, so the intensity left at the screen depends on whether the two waves interfere constructively or destructively there, which can be computed from their wavelengths and the distance of that point of the screen from the slits. Young observed these diffraction and interference patterns, and this work was important for establishing light in classical physics as a wave.

But more puzzling versions of the experiment have been performed starting in the 1920’s. Here are they ordered by increasing puzzling behavior:

- Beams of matter particles, shot at random directions through the slits, also exhibit diffraction and interference patterns, as opposed to leaving on the screen the simple marks expected from a beam of classical particles. That is, the particles behave like waves when they go through the slits and interfere with each other.
- It’s possible to shoot one particle at a time at the slits and gather the statistics of where they hit the screen over time, and the same diffraction and interference patterns show up. Somehow each particle is interfering with itself on its path to the screen.
• In the previous item, if we place a detector in front of the slits in order to know which one the particle actually went through, we no longer find the patterns characteristic of waves on the screen, but rather just the expected classical matter patterns. It’s as if the particle could be in a superposition of two possibilities at the same time (the two slits) until we decide to look and learn which one it was. But if we don’t look we can’t say that it was already following one of the two paths, since in that case it would still interfere with itself. This is related to the collapse of the wavefunction, proposed under the Copenhagen interpretation a few years later.

In 1923, de Broglie proposed a radical idea in his PhD thesis: every particle is actually a wave whose properties can be defined in terms of the things that we think are the energy and momentum of the particle. This would explain the first two items of the list above (though de Broglie was not trying to explain any experiments; these were only starting to be performed around that time).

To start with, a photon has momentum \( p = E/c \). This comes from a relativistic relation that relates energy and momentum for any particle of mass \( m \) (in particular also for a photon, for which \( m = 0 \)):

\[
E^2 = p^2 c^2 + m^2 c^4
\]

If you don’t feel comfortable with relativity, rest assured that the energy-momentum relation of a photon was actually known empirically much before the 1900’s, due to the discovery of radiation pressure. Now plug-in Einstein’s \( E = h \nu \) to get \( p = h \nu / c = h/\lambda \). This led de Broglie to postulate that the wave associated to a particle of momentum \( p \) has wavelength and wave-vector determined by \( p \) as:

\[
\begin{align*}
\lambda &= \frac{h}{p} \\
k &= \frac{p}{\hbar}
\end{align*}
\]

The wave-vector formula comes from the classical formula \( \lambda = 2\pi/k \) and the sensible assumption that \( k \) and \( p \) should be parallel. These are called the de Broglie-Einstein relations, since Einstein’s \( E = h \nu \) is at the root of it (and since Einstein also hinted at them briefly in 1909). A series of experiments by Davisson and Germer, conducted between 1923 and 1927, showed the wave-like behavior of matter and helped establish de Broglie’s theory. Sometimes you will also find the following mentioned among the de Broglie relations:

\[
\omega = \frac{E}{\hbar}
\]

This is coming from \( E = h \nu \) together with \( \omega = 2\pi \nu \).

Remark. Some people like to point out how the de Broglie relations imply that \( \lambda \) is small for objects of large mass (large \( p \)), hence that’s why we don’t see macroscopic objects behaving like waves; their corresponding wavelengths are too small to detect. This type of argument is not rigorous because a macroscopic object is not just one particle. Presumably it’s the interaction of the waves of all the particles in it that somehow accounts for it behaving as we would expect from a classical object.

Remark. Under de Broglie’s interpretation of matter as waves, Bohr’s quantization rule is better understood as stipulating that the wave for an electron orbiting a nucleus should be a standing wave, that is, the orbital circumference should be an integer multiple of its wavelength, so that the wave would look like it’s only oscillating in place, not also moving around the nucleus:

\[
2\pi R = n\lambda \quad , \quad n = 1, 2, 3, \ldots
\]

Indeed, using \( \lambda = \frac{h}{p} = \frac{h}{m\dot{r}} = \frac{h}{m} \sqrt{\frac{mR}{e^2}} = \frac{h\sqrt{R}}{e\sqrt{m}} \), we find

\[
\frac{2\pi R}{n} = \frac{h\sqrt{R}}{e\sqrt{m}} \quad \Rightarrow \quad R = \left( \frac{nh}{2\pi e\sqrt{m}} \right)^2 = \frac{n^2h^2}{me^2} = n^2a_0 \quad , \quad a_0 = \frac{h^2}{me^2}
\]

from which the original Bohr quantization of angular momentum follows as above.
5. Establishment of Quantum Physics

All of the above served to pave the way for a more well-defined theory, developed in the 1920’s, and still in use today: modern quantum mechanics. Let’s talk now about some of its premises and some of the motivation for its postulates. Things covered in this section, unlike the previous one, are considered part of the modern theory.

5.1. Born rule. For a wave of the form

\[ \psi(x, t) = \psi_0 e^{i(kx - \omega t)} \]

the quantity \(|\psi|^2 = |\psi_0|^2\) is a measure of its intensity. In 1927 Born decided to interpret \(|\psi(x, t)|^2\), for a given matter wave \(\psi\), as a function of \(x\) that gives a probability density for finding the corresponding particle at that position at time \(t\). Therefore only \(L^2\)-normalized waves are admitted for representing matter waves:

\[
\int_{\mathbb{R}^3} |\psi(x, t)|^2 d^3x = 1 \quad \text{for all } t
\]

Regions where \(\psi\) is more intense are places where it’s more likely to find the particle, but unless the wave is concentrated (a \(\delta\) measure) there will always be uncertainty about the true value of the position. And mathematically a \(\delta\) measure is not an element of \(L^2\) (there are no functions in \(L^2(\mathbb{R}^3)\) that are zero on all except one point), so that absolute localization of a particle is impossible.

In fact, most people at that time and even today were thinking more on the lines of completely doing away with the notion of a particle being a point that exists in space and whose wavefunction somehow represents just our uncertainty about its true location; instead the only real entity is the wavefunction itself, and when it’s concentrated on microscopic regions of space it looks to us like a point particle sitting there. Matter as we know it, when we decide to inspect closer, is usually composed of many waves of this concentrated form, which we like to think of as particles because we can’t contemplate their tiny extent. But nothing prevents them to become more widespread when we’re not looking and start exhibiting true wave-like behavior.

Now consider \(t = 0\) for simplicity of notation. If \(\psi\) is given as a superposition of wave packets, each having their own wave-vector \(k\):

\[
\psi(x) = \frac{1}{(2\pi)^{3/2}} \int \phi(k) e^{i k x} d^3k
\]

then the Plancherel formula says that \(|\phi|^2\) can also be interpreted as a probability density:

\[
\int_{\mathbb{R}^3} |\phi(k)|^2 d^3k = 1
\]

Born interpreted it as the probability density for the wave-vector \(k\) (after all, that’s what the \(k\) on the exponential represents). Change the variable to the momentum \(p = \hbar k\) associated to each packet (according to de Broglie’s relations) and define

\[
\varphi(p) = \frac{1}{\sqrt{\hbar^3}} \phi \left( \frac{p}{\hbar} \right)
\]

in order to preserve \(L^2\) norm 1. The quantity \(|\varphi(p)|^2\) is then interpreted as the probability density for finding the momentum \(p\). Then the relationship between \(\psi\) and \(\varphi\) is

\[
\psi(x) = \frac{1}{(2\pi\hbar)^{3/2}} \int \varphi(p) e^{i p x} d^3p
\]

This is why we introduced the Fourier transform with parameter before, \(F_h\), which can now be used to obtain \(\varphi\) from \(\psi\):

\[
\varphi(p) = (F_h \psi)(p) = \frac{1}{\sqrt{(2\pi\hbar)^3}} \int_{\mathbb{R}^3} \psi(x) e^{-i p x} d^3x
\]

So \(\psi\) and \(\varphi\) are just two different pictures of the same reality. Only one is necessary to fully describe this “particle” or whatever it is, and the other can be obtained from it through the Fourier transform. What our minds think of as the position and momentum of a particle really are just random variables, each one following its own probability distribution, derived from a true physical entity, the wave-function.
These random variables don’t necessarily have physical meaning, although somehow they correspond to what classically is thought of as position and momentum of a body that moves through space.

5.2. Gaussian Wave Packet. This is the easiest example showing the relation between the two different wavefunctions associated with a particle, \( \psi(x) \) and \( \varphi(p) \). First consider the standard Gaussian in 1D:

\[
f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}
\]

This has total integral equal to 1. We can compute its Fourier transform \( \mathcal{F}_\hbar \) by completing the square and performing a standard change of complex path integral:

\[
(\mathcal{F}_\hbar f)(p) = \frac{1}{\sqrt{2\pi\hbar}} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{-\frac{i}{\hbar}px} dx = \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{i}{\hbar}pk} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - \frac{p}{\hbar})^2} dx = \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{x^2}{2\hbar^2}} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{x^2}{2\hbar^2}} = \frac{1}{\sqrt{\hbar}} f\left(\frac{p}{\hbar}\right)
\]

This has integral equal to \( \sqrt{\hbar} \) (think carefully, this has nothing to do with Plancherel). Using this formula we can deduce the Fourier transform of a translated and rescaled version of \( f \) (to be used shortly): let

\[
g(x) = e^{-\frac{(x-k_0)^2}{4\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}
\]

for some \( k_0 \in \mathbb{R} \) and \( \sigma > 0 \). Then

\[
(\mathcal{F}_\hbar g)(p) = \sqrt{2\pi} e^{-\frac{i}{\hbar}pk_0} \sqrt{2\sigma} (\mathcal{F}_\hbar f)(\sqrt{2\sigma}p) = \sqrt{2\pi} e^{-\frac{i}{\hbar}pk_0} \sqrt{2\sigma} \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{(\sqrt{2\sigma}p\hbar)^2}{2}} = \sqrt{2 \sigma} e^{-\frac{i}{\hbar}pk_0} e^{-\frac{\sigma^2p^2}{\hbar^2}}
\]

Now consider a matter wave \( \psi \) such that \( |\psi|^2 \) is a rescaled Gaussian with total integral 1, such as

\[
\psi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{4\sigma^2}} e^{\frac{i}{\hbar}k_0 x}
\]

Indeed, we have

\[
|\psi(x)|^2 = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)
\]

which has the same integral as \( f \) by the change of variable \( x/\sigma \mapsto x \). This Gaussian has standard deviation \( \sigma \). Its momentum wave function is

\[
\varphi(p) = (\mathcal{F}_\hbar \psi)(p) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathcal{F}_\hbar \left(e^{-\frac{(x-k_0)^2}{4\sigma^2}}\right)(p) = \frac{1}{\sqrt{2\pi\sigma^2}} \sqrt{\frac{2}{\pi}} e^{-\frac{\sigma^2p^2}{2\hbar^2}} e^{-\frac{i}{\hbar}pk_0} = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2\hbar^2}} e^{-\frac{i}{\hbar}pk_0}
\]

and the associated probability density is

\[
|\varphi(p)|^2 = \sqrt{\frac{2\sigma^2}{\pi\hbar^2}} e^{-\frac{\sigma^2p^2}{\hbar^2}}
\]

which is a Gaussian of standard deviation \( \frac{\hbar}{2\sigma} \).

The takeaway is that increasing \( \sigma \) makes this density more localized, but at the cost of making the density for position more spread out, and vice-versa. That is, given the probabilistic interpretation of the functions \( |\psi|^2 \) and \( |\varphi|^2 \), if we want to be able to predict this particle’s momentum to a big degree of accuracy, we won’t be able to know much about its position, and vice-versa. The product of the two standard deviations is \( \frac{\hbar}{2} \), which turns out to be the minimum value possible, as per the next subsection.

5.3. Uncertainty Principle. The calculation below, in this context, was first performed by Heisenberg in 1927. But it’s really consequence of a fact that was probably already known much before: if a function is well localized around a point, its Fourier transform will be more spread out.

Heisenberg’s uncertainty principle. The product of the uncertainties in position and momentum (measured along a given direction) of a matter wave is always at least \( \hbar/2 \):

\[
\sigma_x \sigma_p \geq \frac{\hbar}{2}
\]
Proof. Suppose for simplicity a matter wave \( \psi \) in 1D has zero average for both position and momentum. This is possible by replacing

\[
\psi(x) \mapsto e^{i p_0 x / \hbar} \psi(x - x_0)
\]

where \( x_0 \) and \( p_0 \) are its original averages of position and momentum, because this new \( \psi \) will have zero averages and the exponential in front won’t affect \( |\psi(x)|^2 \).

The standard deviation of position is

\[
\sigma_x = \sqrt{\mathbb{E}[x^2] - \mathbb{E}[x]^2} = \sqrt{\mathbb{E}[x^2]}
\]

where the probability distribution in position space is \( |\psi(x)|^2 \). That is,

\[
\sigma_x = \left( \int_{-\infty}^{\infty} x^2 |\psi(x)|^2 dx \right)^{1/2} = \|x\psi\|_2
\]

Similarly, since \( |\mathcal{F}_\hbar \psi|^2 \) is the probability distribution in momentum space, the standard deviation of momentum is

\[
\sigma_p = \left( \int_{-\infty}^{\infty} p^2 |\mathcal{F}_\hbar \psi(p)|^2 dp \right)^{1/2} = \|p \mathcal{F}_\hbar \psi\|_2
\]

We assume \( \|\psi\|_2 = 1 \). We may also assume (technicalities involving dense subspaces of \( L^2 \)) that \( \psi \) is differentiable and has compact support. This allows us to do integration by parts and simply ignore boundary terms:

\[
1 = \left| \int |\psi|^2 \right| = \left| \int \frac{d}{dx} (x) |\psi|^2 \right| = \left| x |\psi|^2 \right|_\infty - \int x \frac{d}{dx} |\psi|^2 \left. \right| = -2 \Re \int x \bar{\psi} \psi' \leq 2 \left| \int x \bar{\psi} \psi' \right|
\]

Hölder’s inequality says this is

\[
\leq 2 \left( \int |x| |\psi|^2 \right)^{1/2} \left( \int |\psi'|^2 \right)^{1/2} = 2 \sigma_x \|\psi'|\_2
\]

This last integral is treated by the isometry and derivative properties of the Fourier transform:

\[
\|\psi'|\_2 = \|\mathcal{F}_\hbar (\psi')\|_2 = \left| \frac{i p}{\hbar} \mathcal{F}_\hbar \psi \right\|_2 = \frac{1}{\hbar} \sigma_p
\]

Overall we get \( 1 \leq \frac{2}{\hbar} \sigma_x \sigma_p \), so \( \sigma_x \sigma_p \geq \frac{\hbar}{2} \).

In 3D, we have to define random variables \( x_1, x_2, x_3 \) and \( p_1, p_2, p_3 \) which represent the coordinates of position and momentum (which are vectors) in some basis, and then the theorem will say

\[
\sigma_{x_j} \sigma_{p_j} \geq \frac{\hbar}{2}, \quad j = 1, 2, 3
\]

Remark. This provides a good heuristic for why quantum matter, contrary to classical matter, is stable. When we were doing classical physics we saw that an electron orbiting a nucleus could be made to lose energy by going into closer and closer orbits around it. But this implies we’re causing its position to be more and more localized, and in quantum physics this means its momentum necessarily is getting larger and larger. Eventually it gets so large that the electron is not bound to the nucleus anymore and escapes. Of course this is just a layman’s justification. We’ll get more rigorous later.

5.4. Matrix mechanics. In 1925 Heisenberg, Born and Jordan created a new quantum theory based on infinite dimensional matrices. According to them, a property of a physical system (like the energy of an electron orbiting a nucleus) has a discrete set of possible values and there are probabilities of it instantaneously transitioning from one such value to another upon interaction with some external agent (a quantum jump). These probabilities can be encoded as elements of a symmetric matrix. So now the usual laws of classical physics involving for example energy, position, momentum, angular momentum etc. should be written for matrix quantities, and the fact that matrices don’t necessarily commute creates additional terms in the equations which are not present in the classical theory.
This theory never became widely accepted, particularly by the Einstein school of thought, because of the underlying role of discreteness already present within the system since the start. Most physicists were still looking for a continuous theory, and some (Einstein included) believed that the quantum jumps observed in the Bohr model were just emergent phenomena, misinterpreted from an undiscovered continuous theory.

Matrix mechanics gained more attention when Pauli was able to derive the equations for the Hydrogen spectrum using it, in 1926. No Nobel prize was awarded to Heisenberg, Born and Jordan for it, and there's speculation that this had to do with Jordan's affiliation to the Nazi party. Heisenberg soon did important work unrelated to it (uncertainty principle) and got his Nobel, but Born had to wait some decades for his prize, which was awarded for his 1927 probabilistic interpretation of quantum mechanics. Then matrix mechanics quickly died after Schrödinger invented a theory of wave mechanics to supplant it, which was later proved to be equivalent to it. That's the next topic.

5.5. Schrödinger equation. Since matter can be thought of as waves, Schrödinger decided to look for a wave equation that would describe the dynamical laws of a system. He initially failed because he was trying to incorporate relativistic corrections which had recently been added to the Bohr model. But when he published his initial, nonrelativistic equation in 1926 and computed the spectrum of Hydrogen from it, it was well-received by most of the physics community, especially those who opposed Heisenberg's overly abstract matrix theory.

Consider a simple plane-wave with constant speed \( \psi(x, t) = \psi_0 e^{i(kx - \omega t)} \), \( \omega \) constant

Taking into account the de Broglie-Einstein relations (5) and (6), we rewrite this as

\[
\psi(x, t) = \psi_0 e^{\frac{i}{\hbar}(p \cdot x - Et)}
\]

The scalar \( E \) and the vector \( p \) are recoverable from \( \psi \) by differentiating:

\[
\partial_t \psi = -\frac{i}{\hbar} \psi_0 E e^{\frac{i}{\hbar}(p \cdot x - Et)} = -\frac{i}{\hbar} \psi \Rightarrow i\hbar \partial_t \psi = E \psi
\]

\[
\nabla \psi = \frac{i}{\hbar} \psi_0 p e^{\frac{i}{\hbar}(p \cdot x - Et)} = \frac{i}{\hbar} \nabla \psi \Rightarrow -i\hbar \nabla \psi = p \psi \Rightarrow -\hbar^2 \Delta \psi = p^2 \psi
\]

Now, if this wave represents a free particle of mass \( m \) (moving without the influence of any potential energy), its energy is just the kinetic part:

\[
E = \frac{p^2}{2m}
\]

Then we find a first-order diffusion-type equation satisfied by the wave:

\[
\left[ i\hbar \partial_t \psi = -\frac{\hbar^2}{2m} \Delta \psi \right]
\]

This is the free Schrödinger equation. It’s like a heat equation, but with a complex coefficient in front of the time derivative.

We also see that \( \psi \) is an eigenfunction for the differential operators \( i\hbar \partial_t \) and \( -i\hbar \nabla \) (this last one is actually a vector of 3 different operators: \( -i\hbar \partial_x, -i\hbar \partial_y, -i\hbar \partial_z \)), with eigenvalues given by the energy and momentum\(^3\). This seems to suggest that these operators should be associated with the idea of observable energy and momentum, and more generally any observable should correspond to an operator of some sort which can act on a wavefunction \( \psi \), in such a way that a wavefunction that represents a particle having value \( \lambda \) for that observable should be a (generalized) eigenfunction of that operator with eigenvalue \( \lambda \).

Allowing now a potential \( V(x) \) to be present, the equation for energy is

\[
E = \frac{p^2}{2m} + V(x)
\]

and presumably the wave representing a particle moving in this scenario will be a superposition of many simple plane-waves like above with \( \omega = \omega(k) \) (equivalently \( E = E(p) \)), so that it can exhibit different patterns

\(^3\)Here we should be careful and keep in mind that, depending on the Hilbert space where we're working, we may actually be talking about a physicist's vague notion of generalized eigenfunctions and eigenvalues, which is the case when the function is not actually \( L^2 \). See the appendix.
of dispersion depending on what the potential is. It’s actually not very illuminating to apply $i\hbar \partial_t$ and $-i\hbar \nabla$ directly to a packet $\psi_0 e^{\frac{i}{\hbar} (p \cdot x - E(p)t)}$, because of the $p$ dependence of $E$ now. Instead, simply multiply (7) by $\psi$ to get

$$i\hbar \partial_t \psi = -\frac{\hbar^2}{2m} \Delta \psi + V\psi$$

We interpret the right-hand side as an operator (the Hamiltonian) applied to $\psi$:

$$H = -\frac{\hbar^2}{2m} \Delta + \hat{V}$$

where $\hat{V}$ is the operator $\psi(x) \mapsto V(x)\psi(x)$. Then the Schrödinger equation is also written

$$i\hbar \partial_t \psi = H\psi$$

But if $\psi$ is actually going to be a superposition of plane-waves, we should not be able to say it satisfies this equation, which came from assuming it was just one plane-wave. What justifies this reasoning, however, is the “fact” that any function $L^2(\mathbb{R}^3)$ can be written a superposition of eigenvectors for $H$, and those are plane-waves. Since the equation they satisfy is linear, $\psi$ will also satisfy the same equation. The reason why “fact” is between quotes and the way to make this justification rigorous is the spectral theorem of self-adjoint operators (appendix).

In particular we first of all will need our $H$ to be self-adjoint. This is actually a stronger condition than simply symmetric when we deal with unbounded operators, which is the case here. In $L^2$, symmetry of $H$ means

$$\int \overline{\psi} H \phi = \int \overline{H \psi} \phi \quad , \quad \forall \phi \in S$$

where $S \subseteq L^2$ is some subspace of $L^2$ of functions to which $H$ can be applied. For now let’s just check symmetry. Notice how the operators that compose our $H$, that is $i\hbar \partial_t, -i\hbar \nabla, V(x)$, while not definable on the whole $L^2(\mathbb{R}^3)$, can be defined as symmetric operators on suitable subspaces. Symmetry for a first-order linear differential operator (with an $i$ or $-i$ in front) is given by integration by parts, assuming the functions in its domain have compact support so that there is no boundary term:

$$\int_{\mathbb{R}^3} -i\hbar \partial_j (\psi) \phi = \int_{\mathbb{R}^3} \overline{\psi} (-i\hbar \partial_j \phi) \quad , \quad j = x, y, z$$

Symmetry for a multiplication operator by a real-valued function is trivial:

$$\int_{\mathbb{R}^3} V(x) \psi \phi = \int_{\mathbb{R}^3} \overline{\psi} V(x) \phi$$

Now that we have symmetry, self-adjointness turns out to follow without much effort, at least for this simple Hamiltonian. But in general, proving the self-adjointness of the Hamiltonian of a quantum system under study is a major challenge of making the theory mathematically rigorous, one that is often skipped over by less mathematically inclined physicists.

**Remark.** The functional calculus actually provides us with the solution of the Schrödinger equation once the initial state is known:

$$\psi(x, t) = e^{-\frac{i}{\hbar} t H} \psi(x, 0)$$

Functional calculus is a tool that allows us to apply any bounded measurable function defined on $\mathbb{R}$, like $e^{ix}$, to any self-adjoint operator, and obtain a bounded operator that satisfies basically the same properties that the function does. But this expression is more important in proving theorems, less so in computing real-life examples, given the abstract nature of the functional calculus.

**Remark.** The fact that the potential part of the energy relation became a simple multiplication operator suggests that observables related only to position should be associated to the operator of multiplication by their corresponding functions (or vectors of functions) of $x = (x_1, x_2, x_3)$. These operators usually are denoted by the same name of the function, but in uppercase and with a hat. For example, the position of the particle is associated with the vector of 3 operators

$$\psi = \psi(x_1, x_2, x_3) \mapsto \hat{X} \psi = (x_1 \psi(x_1, x_2, x_3), x_2 \psi(x_1, x_2, x_3), x_3 \psi(x_1, x_2, x_3))$$
Momentum, on the other hand, is the differential operator \( \hat{P} = -i\hbar \nabla \). Why is position a simple multiplication operator while the complementary momentum is a totally different differential operator? It is only because the function \( \psi \) is originally assumed to be in the position representation, that is, \( |\psi|^2 \) represents a probability density for position. We could just as well formulate the theory for wavefunctions \( \varphi = \mathcal{F}_h \psi \) of momentum, and then \( \mathbf{p} \) would become the operator \( \hat{P} \) of multiplication by \( \mathbf{p} = (p_1, p_2, p_3) \), with \( \mathbf{X} \) now being a vector of differential operators.

5.6. Hydrogen atom. The way to solve a simple Schrödinger equation as above is by separation of variables. Consider the energy expression for an electron of charge \(-e\) orbiting around a nucleus of charge \(+e\) (a Hydrogen atom):

\[
E = \frac{p^2}{2m} - \frac{e^2}{r}
\]

(Position is usually denoted by \( \mathbf{r} \) instead of \( \mathbf{x} \) in this context). It gives rise to a Hamiltonian

\[
H\psi = -\frac{1}{2m} \Delta \psi - \frac{e^2}{r} \psi
\]

The Schrödinger equation is

\[
\hat{H} \psi = -\frac{\hbar^2}{2m} \Delta \psi - \frac{e^2}{r} \psi
\]

The standard trick to solve this is to take advantage of spherical symmetry and look for a solution that separates in spherical coordinates\(^4\), as well as in the time variable: we look for a solution

\[
\psi(\mathbf{r}, t) = T(t) R(r) \Theta(\theta) \Phi(\phi)
\]

where \( \mathbf{r} = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta) \). First we have to figure out how to write \( \Delta \) in spherical coordinates:

\[
\Delta = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2}
\]

Then the Schrödinger equation, divided by \( \psi \), becomes

\[
\frac{i}{\hbar} \frac{T'}{T} = -\frac{\hbar^2}{2m r^2} \left( \frac{1}{R} \frac{\partial}{\partial r} (r^2 R') + \frac{1}{\Theta \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \Theta') + \frac{1}{\Phi \sin^2 \theta} \Phi'' \right) - \frac{e^2}{r}
\]

Both sides must be equal to some constant \( E \), since all the \( t \)'s are on the left and all the other variables on the right. Then the \( T \) equation is easy to solve:

\[
T(t) = e^{-\frac{i}{\hbar} Et}
\]

This implies that the time-independent part \( \varphi(\mathbf{r}) = R(r) \Theta(\theta) \Phi(\phi) \) satisfies the so-called time-independent Schrödinger equation:

\[
H\varphi = E\varphi
\]

That is, this \( E \) will turn out to be an eigenvalue of \( H \) with \( \varphi \) as eigenfunction.

Setting the right side of (8) equal to \( E \) gives

\[
-\frac{\hbar^2}{R} \frac{d}{dr} (r^2 R') - \frac{\hbar^2}{\Theta \sin \theta} \frac{d}{d\theta} (\sin \theta \Theta') - \frac{\hbar^2}{\Phi \sin^2 \theta} \Phi'' = 2mr^2 \left( E + \frac{e^2}{r} \right)
\]

(there is a good reason for keeping the \(-\hbar^2\) on the left). By the same standard trick with separation of variables, the collection of all the \( \theta, \phi \) terms must be some constant \( \tilde{E} \), so we can write

\[
-\hbar^2 \frac{\sin \theta}{\Theta} \frac{d}{d\theta} (\sin \theta \Theta') - \hbar^2 \frac{\Phi''}{\Phi} = \tilde{E} \sin^2 \theta
\]

Now \( \phi \) only shows up in the second term, so this term too must be some constant \( \tilde{E} \):

\[
-\hbar^2 \Phi'' = \tilde{E}\Phi
\]

\(^4\)Here, as in most of physics literature, \( \theta \) and \( \phi \) are the polar and azimuthal angles, respectively (not what a mathematician is used to).
Since the (azimuthal) angle $\phi$ is $2\pi$-periodic, the only possible solutions are\(^5\)
\[
\Phi(\phi) = e^{i\mu\phi}, \quad \hat{E} = (\mu \hbar)^2, \quad \mu \in \mathbb{Z}
\]
This will be part of the source of discreteness in the eigenvalues of the Hamiltonian: **boundary conditions**, that is, the stipulation that $\Phi(0) = \Phi(2\pi)$.

Going back to (10) with this expression for $\Phi$ transforms that equation into
\[
(12) \quad -\hbar^2 \sin \theta \cdot \partial_\theta (\sin \theta \cdot \partial_\theta (\Theta \Phi)) = (\hat{E} \sin^2 \theta - (\mu \hbar)^2) \Theta \Phi
\]
Its solution is more complicated than (11), involving some simplifying changes of coordinates and familiarity with **Legendre polynomials**.

**COMPUTATION TO BE INCLUDED IN THE FUTURE**

What comes out is that $\hat{E}$ must be of the form $(l(l+1))\hbar^2$ for a nonnegative integer $l \geq |\mu|$, and the solution is denoted
\[
\Theta(\theta)\Phi(\phi) = Y_l^\mu(\theta, \phi)
\]
and called a **spherical harmonic**. These functions are harmonic (zero Laplacian) on the sphere and mutually orthogonal. As it turns out, if we define the vector of **angular momentum** operators
\[
\hat{L} = -i\hbar(\hat{\mathbf{X}} \times \nabla)
\]
and consider the differential operators $\hat{L}_3^2$ (absolute value squared) and $\hat{L}_3$ (angular momentum in the direction of the $z$-axis, which is the one that defines the azimuthal angle), we can calculate and find
\[
\hat{L}_3 = -i\hbar \frac{\partial}{\partial \phi}, \quad \hat{L}_3^2 = -\frac{\hbar^2}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) - \frac{\hbar^2}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2}
\]
This is why we kept $\hbar^2$ on the left in (9). So equation (10) can be thought of as an eigenvalue problem for the spherical Laplacian or for the angular momentum squared. Mathematicians already knew how to solve (12) before it was needed in quantum physics.

Finally, going back to (9), we obtain the $r$ equation:
\[
\frac{1}{R} \frac{d}{dr} (r^2 R') = -\frac{2m_\mathrm{e}^4}{\hbar^2} \left( E + \frac{\mathrm{e}^2}{r} \right) + l(l+1)
\]
This one is even more finicky than the others. It can be transformed into a **Laguerre equation** and solved from there. For each choice of $l \in \mathbb{Z}_{\geq 0}$, we find solutions $R(r) = R_{nl}(r)$ that depend on $l$ and on an integer $n > l$, and the associated value of $E$ depends only on $n$:
\[
E = E_n = -\frac{1}{n^2} \frac{m_\mathrm{e}^4}{2\hbar^2} = -\frac{R_{\text{rad}}}{n^2}
\]
The parameters $n, l, \mu$ are called the **principal**, **azimuthal** and **magnetic** quantum numbers.

Therefore the Schrödinger equation naturally finds the energy levels of the Bohr model, but not by an ad hoc procedure as that model did. It also predicts degeneracy, that is, independent eigenfunctions can be associated to the same energy eigenvalue. Indeed, for each $n$, there are $n$ possible values for $l$, and for each one of those, $2l+1$ values for $\mu$, so that the total number of $Y_l^\mu$ associated to energy $E_n$ is $1 + 3 + 5 + \cdots + n = n^2$. We can also compute the average value of $r$ for the wavefunction $R_{nl}Y_l^\mu$ and see that it will yield the appropriate value from the Bohr model ($n^2a_0$). So now we have a new way to talk about an electron being at a certain orbit around the nucleus, like Bohr did, without the need for an actual point-like electron to be present there: it just means the wavefunction being considered is given by an $R_{nl}Y_l^\mu$ whose energy $E_n$ is the one associated to that orbit. Because of this the wavefunctions are called **orbitals**.

The degeneracy helped explain some features of the periodic table, related to how many electrons can occupy the same orbital. It explains different intensities of different spectral lines, coming from the fact that more electrons are allowed on higher orbits. But more work would be needed to fully explain exactly how many (the Pauli exclusion principle).

---

\(^5\)What we call $\mu$ is conventionally called $m$, but we were already using $m$ for the electron mass.
6. Copenhagen interpretation

Niels Bohr founded a research institute for theoretical physics at the University of Copenhagen in 1921 (later renamed the Niels Bohr Institute). Bohr, Heisenberg, Jordan, Pauli and many of the other founding fathers of QM would all eventually work there or at least correspond with people from there in that decade, and today what is known as the Copenhagen interpretation of QM is a somewhat vague set of ideas developed mostly by them around this time, ideas that should be followed by any quantum description of a physical system, and that are still considered true today (not without controversy). They can be described as:

- A physical system is described by a wavefunction $\psi(x, t)$, also called a state. The Hilbert space $\mathcal{H}$ where $\psi$ lives will depend on how many particles or how many properties are needed to specify to fully describe the system. A typical example is $\mathcal{H} = L^2(\mathbb{R}^{3N})$ for a system of $N$ particles. So we no longer have one wavefunction for each particle, we just have one wavefunction for the entire system.

- A Hamiltonian self-adjoint operator $H$, defined on some subspace of $\mathcal{H}$, is associated to the energy of the system. The physical laws obeyed by the system are all encoded in $H$, since the evolution of the state in time is given by the Schrödinger equation $i\hbar \partial_t \psi = H \psi$. This equation is obeyed at almost all times (see collapse of the wavefunction below).

- Nature is not deterministic, but rather probabilistic. The evolution of the wavefunction is deterministic, but its manifestation in the physical universe happens though the function $|\psi|^2$, which is used as a probability density with respect to which probabilistic statements about the system can be asserted. The Schrödinger equation can be used to easily show that, if $\psi(x, t)$ satisfies $\|\psi(x, 0)\|_2 = 1$, then it automatically satisfies $\|\psi(x, t)\|_2 = 1$ for all $t$. Simply compute $\partial_t \langle \psi, \psi \rangle$ using symmetry of $H$ and check it’s zero.

- Superposition: Sometimes a quantum system may be in a state of superposition of two or more particular states. This is achieved mathematically by saying its wavefunction is the sum of the corresponding wavefunctions of those states (properly normalized).

- Correspondence principle: Laws of large numbers should be in place to ensure that the quantum mechanical description of large systems will closely approximate the classical description. This is why macroscopic, everyday physics doesn’t seem probabilistic.

- Uncertainty principle: Certain properties of a state can never be jointly determined with arbitrary precision. This is not stemming from inaccuracies in the measuring device, it’s an inherent mathematical property of the wavefunctions.

- An observable is a property of the system that (usually) has a classical counterpart and is encoded by a self-adjoint operator. Examples are the energy, the position and momentum of a particle, the spin of an electron.

- The measurement of an observable is a process that can only be described using language from classical physics. A classical measuring device interacts with the wavefunction and yields a random real number in the spectrum of the corresponding operator. The randomness depends on the state of the system before the measurement.

- Collapse of the wavefunction: When a measurement is performed on the system, its wavefunction instantaneously becomes a (generalized) eigenfunction corresponding to the random (generalized) eigenvalue yielded by the measurement. This is the only time when a state does not follow a deterministic time evolution, like what happens when we decide to try and detect the slit through which an electron is going through in the double-slit experiment.

To be more quantitative, given two states $\psi, \varphi$, the complex number $\langle \varphi, \psi \rangle$ is the coefficient of the orthogonal projection of $\psi$ onto $\varphi$, since $\|\varphi\|^2 = 1$. Its absolute value squared, $|\langle \varphi, \psi \rangle|^2$, is given the vague interpretation of the probability of $\psi$ turning into $\varphi$. Given also an observable $A$, we interpret $|\langle \varphi, A\psi \rangle|$ as the probability that the $\psi$ wavefunction will collapse into $\varphi$ upon being measured by the operator $A$. This allows us to define the average value or expected value of the observable $A$ when the system is in the state $\psi$:

$$\langle A \rangle_\psi = \langle \psi, A\psi \rangle$$
We can also use just $\langle A \rangle$ if $\psi$ is clear from context. A physicist justifies this definition by expanding $\psi = \sum_j \langle \psi_j, \psi \rangle \psi_j$ into a sum over eigenfunctions $\psi_j$ of $A$:

$$A \psi_j = \lambda_j \psi_j$$

which gives

$$\langle \psi, A \psi \rangle = \left\langle \sum_j \langle \psi_j, \psi \rangle \psi_j , \sum_k \langle \psi_k, \psi \rangle A \psi_k \right\rangle$$

$$= \sum_j \sum_k \langle \psi, \psi_j \rangle \langle \psi_k, \psi \rangle \langle \psi_j, A \psi_k \rangle$$

$$= \sum_j \sum_k \lambda_k \langle \psi, \psi_j \rangle \langle \psi_k, \psi \rangle \langle \psi_j, \psi_k \rangle$$

$$= \sum_j \lambda_j |\langle \psi_j, \psi \rangle|^2$$

This is a sum over all possible values of the measurement (the $\lambda_j$’s), each weighted with its probability (the probability $|\langle \psi_j, \psi \rangle|^2$ of the wavefunction $\psi$ collapsing into the eigenstate $\psi_j$), which is the mathematical definition of the expected value. A mathematician, on the other hand, justifies the definition of the expected value by going into an equivalent Hilbert space $L^2(X)$ where $A$ is given as a multiplication operator

$$\psi(x) \mapsto f(x)\psi(x)$$

since then

$$\langle \psi, A \psi \rangle = \int_X \overline{\psi} f \psi = \int_X f |\psi|^2$$

If we remember that $|\psi|^2$ is a probability density, then this is the expected value of the function $f(x)$, whose range is the spectrum of $A$ (the set of all possible measurement results).

The Copenhagen interpretation has some problems which, almost a century later, are not even close to being resolved:

- What constitutes a measurement? Why should the observer have the power to temporarily switch off the deterministic time-evolution of the wavefunction and make it collapse?
- Is it really necessary to keep the classical physics language? One still talks about particles and their position and momentum, even though those are not part of the axioms.
- Is there a way to consider separate wavefunctions for each part of a system? This is frequently employed in both theory and practice, even though there should be only one universal $\psi$.
- What is the importance of self-adjointness? Often the observables being studied are not known to be self-adjoint or, worse, are known not to be, but still the theoretical results may predict things that will actually match the numbers yielded by experiments, so that some other yet unknown mathematical structure is probably hiding in the theory and will be needed if we ever achieve a complete and consistent version of it.

Alternative viewpoints claim to solve these fully or partially, but these claims are still very controversial. We can roughly describe some of them as follows:

- **Decoherence**: To produce a result, a measurement device must interact with the system in such a strong way that the quantum properties of the system “decohere” away into the device and the environment. The evolution of the universe’s wavefunction is still unitary (that is, given by the Schrödinger equation), but only when we consider the environment as part of the system now. It’s only from the point of view of the observer within the system that the illusory wavefunction of the experiment will seem to have suffered collapse.

- **Many-worlds interpretation**: When collapse seems to happen, we are actually just following a branch of the tree of all infinitely many possible time-evolutions from that point on. Each branch is equally real and realized “somewhere else”, in another universe which will never have any relation to our branch ever again. This interpretation thinks of the collapse as an illusion caused, like in the above item, by the fact that the observer is now part of the system with the experiment, but it gives
physical meaning to each possible result. And those results include what the observer saw: they could be for example system has property $A$ and observer measured $A$, system has property $A$ but observer measured property $B$, system has property $B$ and observer measured $B$, system has property $B$ but observer measured $A$.

**de Broglie-Bohm theory:** Point particles are actually real and they have well-defined positions and momenta. The wavefunction is also real, its evolution through the Schrödinger equation serving as a guiding term that enters in a second fundamental equation, one that is coupled to Schrödinger and actually describes the particles’ motion in configuration space, in a manner analogous to the so-called Hamilton-Jacobi formulation of classical mechanics. We can actually talk about a particle’s individual wavefunction by considering the universal $\psi$ conditioned on the actual positions of everything else in the universe.

6.1. **Ehrenfest Theorem.** This theorem is a first step towards understanding how the correspondence principle arises from quantum theory. First of all, given two operators $A, B$, we define their **commutator** and **anticommutator** as the operators

$$[A, B] = AB - BA, \quad \{A, B\} = AB + BA$$

wherever defined. When $A, B$ are Hermitian $^6$, the commutator $[A, B]$ is anti-Hermitian (that is, $[A, B]^* = -[A, B]$), as can be easily checked.

**Lemma.** Let $A$ be an observable, which may be time-dependent. Then

$$\frac{d\langle A \rangle}{dt} = -\frac{i}{\hbar}[A, H] + \langle \frac{\partial A}{\partial t} \rangle$$

The averages refer to any normalized state $\psi$.

**Proof.** Just calculate:

$$\frac{d}{dt} \langle \psi, A\psi \rangle = \langle \psi_t, A\psi \rangle + \langle \psi, A\psi_t \rangle = \frac{1}{i\hbar} H\psi, A\psi \rangle + \langle \psi, A\psi_t \rangle = \langle \psi, (-A^*H + AH)\psi \rangle + \langle \psi, A\psi_t \rangle = -i\hbar \langle [\psi, [A, H]]\psi \rangle + \langle \psi, A\psi_t \rangle$$

In the first equality we had to know how to differentiate $A\psi$ when both $A$ and $\psi$ depend on time, and the answer was just like the product rule. To see why, simply write

$$A(t)\psi(t) = (A(t_0) + \partial_t A(t_0)(t - t_0) + o(t))(\psi(t_0) + \partial_t \psi(t_0)(t - t_0) + o(t))$$

and distribute by linearity to find that $\partial_t A\psi = (\partial_t A)\psi + A\partial_t \psi$.

Now consider a Hamiltonian of the form

$$H = \frac{1}{2m}\hat{p}^2 + V(\hat{x})$$

in 1D for simplicity of notation. Let’s use the theorem to compute the averages of position and momentum:

$$\frac{d\langle \hat{x} \rangle}{dt} = -\frac{i}{\hbar} \left[\hat{x}, \frac{1}{2m}\hat{p}^2\right], \quad \frac{d\langle \hat{p} \rangle}{dt} = -\frac{i}{\hbar} \langle [\hat{p}, \hat{V}] \rangle$$

(since $\hat{x}$ commutes with $\hat{V}$ and $\hat{p}$ commutes with $\hat{p}^2$).

$$\hat{x}\hat{p}^2\psi - \hat{p}^2\hat{x}\psi = x(-i\hbar)^2\psi'' - (-i\hbar)^2\frac{\partial^2}{\partial x^2}(x\psi) = -\hbar^2 x\psi'' + \hbar^2 x'\psi' + \hbar^2 x''\psi = 2\hbar^2 x\psi' = 2i\hbar(-i\hbar x'\psi) = 2i\hbar\hat{p}\psi \Rightarrow [\hat{x}, \hat{p}^2] = 2i\hbar$$

$$\hat{p}\hat{V}\psi - \hat{V}\hat{p}\psi = -i\hbar(\hat{V}\psi)' - (-i\hbar)\hat{V}\psi' = -i\hbar\hat{V}'\psi$$

$^6$**Hermitian** and **self-adjoint** are synonyms, but each term is more common in certain contexts.
Then

\[
\frac{d\langle \hat{X} \rangle}{dt} = -\frac{i}{2m\hbar} \langle 2i\hbar\hat{P} \rangle = \frac{\langle \hat{P} \rangle}{m}, \quad \frac{d\langle \hat{P} \rangle}{dt} = -\frac{i}{\hbar} \langle -i\hbar\hat{V}' \rangle = -\langle \hat{V}' \rangle
\]

In 3D we get \( \nabla \hat{V} \) instead of \( \hat{V}' \). This gives regular Newton’s Laws for the averaged quantities \( \langle \hat{X} \rangle \) and \( \langle \hat{P} \rangle \):

\[
\langle \hat{P} \rangle = m \frac{d\langle \hat{X} \rangle}{dt}, \quad m \frac{d^2\langle \hat{X} \rangle}{dt^2} = \frac{d\langle \hat{P} \rangle}{dt} = -\langle \nabla \hat{V} \rangle
\]

These two equations are known as Ehrenfest’s Theorem. They say that a large system of particles that individually obey the Schrödinger equation for the Hamiltonian (6.1) will on the large behave like a classical system, since the laws obtained for the time evolution of the averaged quantities are just the classical Newton laws. Individual particles may deviate from classical behavior, but the average of them all will be very close to it by the law of large numbers, so that a big ball of many particles will obey classical physics very closely. Of course this doesn’t account for interactions between the particles of this ball; this is just an easy example of how the correspondence principle arises from quantum laws.

6.2. Uncertainty principle, algebraic version. There is a simple algebraic computation that shows the inherent uncertainty of measurements in QM, not only for position and momentum, but actually for most pairs of observables. It relates the inability of accurately measuring both to the fact that they may not commute as operators acting on wavefunctions.

For a Hermitian or anti-Hermitian operator \( A \), define the operator (dependent on the state \( \psi \))

\[
\Delta_A = A - \langle A \rangle I
\]

and the uncertainty

\[
\delta_A = \sqrt{\langle \Delta_A^2 \rangle} = \sqrt{\langle A^2 \rangle - \langle A \rangle^2}
\]

The second equality comes from

\[
\Delta_A^2 = A^2 - 2\langle A \rangle A + \langle A \rangle^2 I \implies \langle \Delta_A^2 \rangle = \langle A^2 \rangle - 2\langle A \rangle^2 + \langle A \rangle^2 = \langle A^2 \rangle - \langle A \rangle^2
\]

**Theorem (Abstract Uncertainty Inequality).** Given two observables \( A, B \), we have

\[
\delta_A \delta_B \geq \frac{1}{2} |\langle [A, B] \rangle|
\]

for any normalized state \( \psi \).

**Proof.** Define \( \chi = \Delta_A \psi \) and \( \phi = \Delta_B \psi \). Since \( \Delta_A, \Delta_B \) are Hermitian,

\[
\langle \chi, \chi \rangle = \langle \Delta_A \psi, \Delta_A \psi \rangle = \langle \psi, \Delta_A^2 \psi \rangle = \langle \Delta_A^2 \rangle
\]

\[
\langle \phi, \phi \rangle = \langle \Delta_B \psi, \Delta_B \psi \rangle = \langle \psi, \Delta_B^2 \psi \rangle = \langle \Delta_B^2 \rangle
\]

\[
\langle \chi, \phi \rangle = \langle \Delta_A \psi, \Delta_B \psi \rangle = \langle \psi, \Delta_A \Delta_B \psi \rangle = \langle \Delta_A \Delta_B \rangle
\]

so Cauchy-Schwarz gives

(13)

\[
|\langle \Delta_A \Delta_B \rangle|^2 \leq \langle \Delta_A^2 \rangle \langle \Delta_B^2 \rangle
\]

Rewrite \( \Delta_A \Delta_B = \frac{1}{2} \{ \Delta_A, \Delta_B \} + \frac{1}{2} [\Delta_A, \Delta_B] \), which decomposes it as a sum of a Hermitian and an anti-Hermitian operator. The anti-Hermitian part \( \frac{1}{2} [\Delta_A, \Delta_B] \) simplifies to \( \frac{1}{2} [A, B] \). Because the expected value of a Hermitian (resp. an anti-Hermitian) operator is real (resp. complex), we get

(14)

\[
|\langle \Delta_A, \Delta_B \rangle|^2 = \frac{1}{4} \left| \langle [A, B] \rangle \right|^2 + \frac{1}{4} \left| \langle [A, B] \rangle \right|^2 + \frac{1}{4} \left| \langle [A, B] \rangle \right|^2 \geq \frac{1}{4} \left| \langle [A, B] \rangle \right|^2
\]

Putting together (13) and (14),

\[
\langle \Delta_A^2 \rangle \langle \Delta_B^2 \rangle \geq \frac{1}{4} \left| \langle [A, B] \rangle \right|^2
\]

■
The most obvious example of a pair of noncommuting operators are position and momentum. Consider again 1D position \( \hat{X} \) and momentum \( \hat{P} \) for simplicity of notation. We have to compute

\[
[\hat{X}, \hat{P}] \psi = x(-i\hbar)\psi' - (-i\hbar)\frac{\partial}{\partial x}(x\psi) = i\hbar \psi \implies [\hat{X}, \hat{P}] = i\hbar I
\]

where \( I \) is the identity operator. Then the theorem says

\[
\delta_{\hat{X}} \delta_{\hat{P}} \geq \frac{1}{2} \langle |i\hbar I| \rangle = \hbar
\]

which is Heisenberg’s uncertainty principle again, this time with an apparently different notion of the uncertainty, which last time we defined as the standard deviation of the probability distribution. You can easily check that they are the same concept; it’s just that this time the emphasis is on the operator, not the wavefunction.

6.3. **Pauli exclusion principle.** [TO BE EXPANDED]

We never talked about spin in these notes. Spin is like an intrinsic angular momentum carried by a particle, almost as if the particle could be a tiny ball that rotates around itself (this is what people initially thought it was). It was first observed in the Stern-Gerlach experiment (1922), which showed that matter particles get deflected by magnetic fields.

The first to give spin a correct mathematical treatment was Pauli in 1927. He thought of it as a fourth quantum number associated to a particle, having only finitely many possible values. All particles fall into one of two categories: bosons (particles having integer spin, thought of as carriers of force, like photons and gluons) and fermions (particles having half-integer spin, thought of as massive matter, like protons, electrons and neutrons). Pauli explained the discrepancy between the observed chemical orbitals and those predicted by the Schrödinger equation by stipulating that no two fermions can share the same set of quantum numbers. Therefore at most two electrons could occupy the same orbital (one with spin up, the other with spin down).

A later theorem of his relates this language to the wavefunction, and is known as the **spin-statistics theorem**. It says that the wavefunction of a system of identical fermions must be antisymmetric with respect to exchanging arguments corresponding to the positions of any two particles.

We can give an intuitive reason of why this must be true. First, since \( |\psi|^2 \) is the probability density for finding the collection of the particles at specific locations, this function must be *symmetric* with respect to exchanging any two particle positions, and that is certainly achieved with an *antisymmetric* \( \psi \), given that \( |\psi|^2 = \overline{\psi}\psi \). But antisymmetry also imposes that \( \psi = 0 \) (therefore also \( |\psi|^2 = 0 \)) at any point where two of its particle position arguments are the same. That is, antisymmetry guarantees that there is a probability 0 of any two particles occupying the same space, which really seems to be a property of matter in the real world (but not necessarily particles that carry force).

7. **Stability of Matter**

[TO BE ADDED]

This section will show the computation of ground state energy (lowest point on the spectrum) of an atom by finding the minimum of \( \langle \psi, H\psi \rangle \) over appropriate spaces of \( \psi \) functions. Theorems about how this energy scales with increasing number of electrons, and how the Pauli exclusion principle is needed to obtain the correct scaling law.

8. **Appendix: Operator Theory**

A Hilbert space \( \mathcal{H} \) is a \( \mathbb{C} \)-vector space (we assume infinite-dimensional and separable) with an *inner product* \( \langle \cdot, \cdot \rangle \) satisfying:

- \( \langle x, x \rangle \geq 0 \), with equality if and only if \( x = 0 \)
- \( \langle z, x + \lambda y \rangle = \langle z, x \rangle + \lambda \langle z, y \rangle \) for \( x, y, z \in \mathcal{H} \) and \( \lambda \in \mathbb{C} \)
- \( \langle x, y \rangle = \overline{\langle y, x \rangle} \) (this implies \( \langle \lambda x, y \rangle = \overline{\lambda} \langle x, y \rangle \)
The norm of an \( x \in \mathcal{H} \) is \( \|x\| = \sqrt{\langle x, x \rangle} \). These properties imply the Cauchy-Schwarz inequality

\[
|\langle x, y \rangle| \leq \|x\| \|y\|
\]

The typical example of a Hilbert space is \( L^2(X) \) for \( X \) a measure space, with the inner product and norm

\[
\langle \psi, \varphi \rangle = \int_X \overline{\psi} \varphi, \quad \|\psi\|^2 = \int_X |\psi|^2
\]

Every Hilbert space admits orthonormal bases, which are countable sets \( \{x_n\} \subseteq \mathcal{H} \) such that
- \( \|x_n\| = 1 \) for all \( n \)
- \( \langle x_n, x_m \rangle = 0 \) if \( n \neq m \)
- Every \( x \in \mathcal{H} \) can be expressed as \( x = \sum_{n=1}^{\infty} \langle x_n, x \rangle x_n \) (the norm is used to give meaning to an infinite sum)

The Parseval identity follows from this:

\[
\|x\|^2 = \sum_{n=1}^{\infty} |\langle x_n, x \rangle|^2, \quad \text{for all } x \in \mathcal{H}
\]

More generally,

\[
\langle y_1, y_2 \rangle = \sum_{n=1}^{\infty} \langle y_1, x_n \rangle \langle x_n, y_2 \rangle, \quad \text{for all } y_1, y_2 \in \mathcal{H}
\]

(Simply expand \( y_1, y_2 \) as infinite series and use \( \langle x_n, x_m \rangle = \delta_{mn} \).

An operator \( T : \mathcal{H} \to \mathcal{H} \) (always assumed linear) is bounded (which means continuous) if and only if there is \( C \geq 0 \) such that

\[
\|Tx\| \leq C\|x\|, \quad \text{for all } x \in \mathcal{H}
\]

The closed graph theorem says that this is equivalent to \( T \) having a closed graph, which means that, in order to prove an operator \( T \) is bounded, we don’t only assume \( x_n \to x \) and prove \( Tx_n \to Tx \); we are also allowed to assume \( Tx_n \to y \) for some \( y \in \mathcal{H} \) and then we must prove \( Tx = y \).

An eigenvalue of an operator \( T \) is \( \lambda \in \mathbb{C} \) for which there exists \( x \in \mathcal{H}, x \neq 0 \), such that \( Tx = \lambda x \). Such an \( x \) is called an eigenvector, or also an eigenfunction when the Hilbert space is a space of functions like \( L^2 \).

The spectrum of the operator \( T \) is

\[
\sigma(T) = \{ \lambda \in \mathbb{C} ; T - \lambda I \text{ does not have a bounded inverse} \}
\]

This is always nonempty and closed. Boundedness of \( \sigma(T) \) is equivalent to boundedness of \( T \). Every eigenvalue of \( T \) is in the spectrum, but there could be other points in it as well (the operator \( T - \lambda I \) may fail to be invertible even if it is injective).

Every bounded operator has an adjoint \( T^* : \mathcal{H} \to \mathcal{H} \), which is the only operator such that

\[
\langle Tx, y \rangle = \langle x, T^* y \rangle, \quad \text{for all } x, y \in \mathcal{H}
\]

\( T \) is called self-adjoint when \( T^* = T \). In this case, only real numbers can be eigenvalues for it, because if \( Tx = (a + bi)x \) with \( b \neq 0 \) then

\[
(a + bi)\langle x, x \rangle = \langle x, (a + bi)x \rangle = \langle x, Tx \rangle = \langle Tx, x \rangle = \langle (a + bi)x, x \rangle = (a - bi)\langle x, x \rangle \Rightarrow x = 0
\]

Without much more effort we can also show that not only the set of eigenvalues, but its entire spectrum is contained in \( \mathbb{R} \).

An orthogonal projection or simply projection is a self-adjoint operator \( P \) satisfying \( P^2 = P \). In that case there is a decomposition \( \mathcal{H} = \ker P \oplus \operatorname{ran} P \), and \( P \) is simply the operator that finds the \( \operatorname{ran} P \) component of any given vector. The spectrum of a (nontrivial) projection is \( \{0, 1\} \).

The spectral theorem says that every self-adjoint operator \( T \) is diagonalizable by a real operator, which means that there exists an isometric (norm-preserving) isomorphism \( \mathcal{H} \to L^2(X) \) for some measure space \( X \) such that \( T \), seen as an operator on \( L^2(X) \), is given as a multiplication operator

\[
L^2(X) \ni \psi(x) \mapsto f(x)\psi(x) \in L^2(X)
\]
for some real-valued bounded measurable function \( f(x) \). The spectrum of such an operator is the range of \( f(x) \) (technically it’s the essential range, that is, the set of \( z \in \mathbb{C} \) such that, for all \( \varepsilon > 0 \), the inverse image \( f^{-1}(B(z,\varepsilon)) \) has positive measure). Isolated points in the spectrum of self-adjoint operators are eigenvalues, but there might also be a continuous part of the spectrum, for example an entire interval contained in it. And sometimes there are no eigenvalues, only continuous spectrum.

**Theorem (Hellinger-Toeplitz).** An operator satisfying \( \langle Tx, y \rangle = \langle x, Ty \rangle \) for all \( x, y \in \mathcal{H} \) is automatically bounded.

**Proof.** Suppose \( x_n \to x \) and \( Tx_n \to y \). Then, for all \( z \in \mathcal{H} \),

\[
\langle z, Tx \rangle = \langle Tz, x \rangle = \lim \langle Tz, x_n \rangle = \lim \langle z, Tx_n \rangle = \langle z, y \rangle
\]

which proves \( Tx = y \).

The significance of this theorem is that, if we ever have to deal with unbounded symmetric operators \( T \) (and they are plentiful in QM), they’ll necessarily not be definable on the entire Hilbert space \( \mathcal{H} \), only on some subspace \( \mathcal{D}(T) \subseteq \mathcal{H} \), which creates some annoyances. The operators that people work with always have at least \( \mathcal{D}(T) \subseteq \mathcal{H} \) dense (otherwise simply restrict your Hilbert space to make that happen). We also generally assume that the graph of \( T \) is closed in \( \mathcal{H} \times \mathcal{H} \), which would give boundedness of \( T \) if it were globally defined on \( \mathcal{H} \), but is nevertheless still a useful property.

For unbounded operators there is a distinction between symmetry (a weaker property) and self-adjointness (much stronger). An operator \( T : \mathcal{D}(T) \to \mathcal{H} \) is symmetric if

\[
\langle Tx, y \rangle = \langle x, Ty \rangle \quad \text{for all } x, y \in \mathcal{D}(T).
\]

In that case its adjoint domain is the set \( \mathcal{D}(T^*) \supseteq \mathcal{D}(T) \) of all \( x \in \mathcal{H} \) for which there exists \( y \in \mathcal{H} \) such that

\[
\langle x, Tz \rangle = \langle y, z \rangle \quad \text{for all } z \in \mathcal{D}(T).
\]

Then the adjoint \( T^* \) is defined on \( \mathcal{D}(T^*) \) by \( T^*x = y \). The operator is said to be self-adjoint when \( \mathcal{D}(T^*) = \mathcal{D}(T) \). Therefore the adjoint of a symmetric operator is an extension of it, and if we want to create a self-adjoint operator from a symmetric one, it will have to be by extending it to a suitable domain. But it turns out that not all symmetric operators have self-adjoint extensions, and some others have uncountable families of them.

The spectral theorem extends to the unbounded case: for every self-adjoint operator \( T \), there exists an isometric (norm-preserving) isomorphism \( \mathcal{H} \to L^2(X) \) for some measure space \( X \) such that \( T \), seen as an operator on \( L^2(X) \), is given as a multiplication operator

\[
L^2(X) \ni \psi(x) \mapsto f(x) \psi(x) \in L^2(X)
\]

for some real-valued measurable function \( f(x) \). But this time nothing guarantees this multiplication operator is defined globally on \( L^2(X) \), since \( f(x) \) may not be bounded. In fact, \( \mathcal{D}(T) \) gets mapped by the isometry between \( \mathcal{H} \) and \( L^2(X) \) into \( \{ \psi \in L^2(X) : f(x) \psi(x) \in L^2(X) \} \).

This permits us to construct a Borel functional calculus of unbounded operators, that is, to make sense of expressions like \( g(T) \) for \( T \) an unbounded operator and \( g \) a bounded Borel-measurable function defined on its spectrum: simply consider the operator given by multiplication with \( g(f(x)) \). Using this we can find, for example, the square-root \( \sqrt{T} \) of an operator with nonnegative spectrum. But most importantly for quantum mechanics, it also allows us to define bounded operators \( T(t) = e^{itT} \), for \( t \in \mathbb{R} \). These satisfy properties that allow us to label them a unitary one-parameter group:

- \( T(0) = I \)
- \( T(t+s) = T(t)T(s) \)
- \( \lim_{t \to 0} T(t)x = T(t_0)x \) for all \( x \in \mathcal{H} \)
- \( T_t^* = T_{-t}^* \)

This concept turns out to be the central notion needed for explaining the time evolution of a quantum mechanical system. Then there’s a result called Stone’s theorem that says that any unitary one-parameter group must arise in the form \( e^{itT} \) for some self-adjoint operator \( T \). So this provides very compelling evidence that self-adjointness is central to quantum mechanics.
The minimax theorem, familiar from matrix theory, has an analogous formulation for unbounded self-adjoint operators. For Hermitian matrices it allows one to find the eigenvalues by finding minima and maxima of Rayleigh quotients over appropriate subspaces. For operators, one of its consequences is that, for any self-adjoint $T$,
\[-\infty \leq \inf \sigma(T) = \inf_{\|x\| = 1} \langle x, Tx \rangle, \quad \infty \geq \sup \sigma(T) = \sup_{\|x\| = 1} \langle x, Tx \rangle\]

Still on the topic of self-adjointness, in the physics literature it’s common to find the following:

**Theorem (diagonalization of self-adjoint operators).** Every self-adjoint operator provides an orthonormal basis of eigenvectors for the underlying Hilbert space.

This is a false statement. There is a way to generalize the notion of an orthonormal basis of eigenvectors to make this true, although it’s more like a heuristic and not really a formal definition. The spectral theorem is the key: every self-adjoint operator $T$ can be thought of as
\[L^2(X) \ni \psi(x) \mapsto f(x)\psi(x) \in L^2(X)\]
for some real-valued measurable function $f(x)$ and some measure space $X$. Its spectrum is the range of $f$.

Proceeding informally for now, take a huge sample of values $\lambda_j, j = 1, \cdots, N$, in ran($f$) and let’s imagine dividing up $X$ into tiny pieces $E_j = \{x \in X : f(x)$ in some sense is close to $\lambda_j\}$, in such a way that each $E_j$ has measure $M_j > 0$ and their disjoint union is $X$. Define the normalized functions $\psi_j \in L^2(X)$ by
\[\psi_j = \frac{1}{\sqrt{M_j}} \chi_{E_j} = \begin{cases} 
\frac{1}{\sqrt{M_j}}, & \text{if } x \in E_j, \\
0, & \text{if not.} 
\end{cases}\]

These are orthogonal to each other by disjointness of the $E_j$, and are approximate eigenfunctions of $T$:
\[T\psi_j = \frac{1}{\sqrt{M_j}} f(x)\chi_{E_j} \approx \frac{1}{\sqrt{M_j}} \lambda_j \chi_{E_j} = \lambda_j \psi_j\]

Moreover, they can be considered as some sort of approximate orthonormal basis, since every $\psi \in L^2(X)$ can basically be written in terms of a sum involving its average value in each of the tiny sets $E_j$, which can be turned into the familiar expression for a vector in an orthonormal basis:
\[\psi \approx \sum_{j=1}^N \left( \frac{1}{M_j} \int_{E_j} \psi \right) \chi_{E_j} = \sum_{j=1}^N \frac{1}{\sqrt{M_j}} \left( \int_X \chi_{E_j} \psi \right) \psi_j = \sum_{j=1}^N \langle \psi_j, \psi \rangle \psi_j. \tag{16}\]

In particular what comes out of this is that $T$ is approximately given as a linear combination of projections $P_j : \psi \mapsto \psi \chi_{E_j}$, with the eigenvalues $\lambda_j$ as coefficients, since for each $\psi$
\[T\psi \approx \sum_{j=1}^N \langle \psi_j, \psi \rangle T\psi_j = \sum_{j=1}^N \left( \int_X \psi_j \psi \right) \lambda_j \psi_j = \sum_{j=1}^N \lambda_j \left( \frac{1}{M_j} \int_{E_j} \psi \right) \chi_{E_j} \approx \sum_{j=1}^N \lambda_j \psi \chi_{E_j} = \sum_{j=1}^N \lambda_j P_j \psi\]
\[\implies T \approx \sum_{j=1}^N \lambda_j P_j \tag{17}\]

Rigorously speaking now, notice that for every Borel set $E \subseteq \text{ran}(f)$ there is associated a projection on $L^2(X)$ given by multiplication by $\chi_{f^{-1}(E)}$. That is, this projection takes any $\psi$ and restricts it to only those points $x$ with $f(x) \in E$. If the measure of the set of those points is zero in $X$, this is just the zero projection, but in general, even if there is only one point $x_0$ such that $f(x_0) \in E$, it might still produce a nontrivial projection if the measure of $\{x_0\}$ is nonzero. What can be proved then is that one can define a projection-valued measure $dP$ on the Borel $\sigma$-algebra of $\text{ran}(f)$, called the spectral measure of $T$. This has the same properties as a standard measure, but the measure of a set is not a number, it’s a projection on
With it, one can create operators on $L^2(X)$ that are given by integrating bounded Borel measurable functions over $\text{ran}(f)$. Such operators act on a $\psi \in L^2(X)$ by multiplication:

$$\left( \int_{\text{ran}(f)} g(\lambda) \, dP(\lambda) \right) \psi = (g \circ f)\psi$$

To get an idea why this is so, notice how, on the left side, each infinitesimal $dP(\lambda)$ can be thought of as a projection that, when acting on $\psi$, produces $\psi$ restricted to just those $x$ for which $f(x) = \lambda$. Then we’re multiplying that by $g(\lambda) = g(f(x))$ at each of those points. Finally, this is done “continuously many” times, and when everything is added up, we get the function $g(f(x))\psi(x)$. In particular the operator $T$ itself is obtained by $g(\lambda) = \lambda$:

$$(18) \quad T = \int_{\text{ran}(f)} \lambda \, dP(\lambda)$$

which generalizes (17), while any function $\psi \in L^2(X)$ can be obtained from the identity operator (let $g(\lambda) = 1$):

$$(19) \quad \psi = \left( \int_{\text{ran}(f)} dP(\lambda) \right) \psi$$

which generalizes (16).

Physicists unknowingly regard a function like $\psi_{\lambda_0} = \chi_{f^{-1}(\lambda_0)}$, which might very well be just the zero function in $L^2$ due to measure considerations, as an eigenfunction of $T$ with eigenvalue $\lambda_0$: $T(\chi_{f^{-1}(\lambda_0)}) = \int_{\text{ran}(f)} \lambda \, dP(\lambda)\chi_{f^{-1}(\lambda_0)} = \int_{\{\lambda_0\}} \lambda \, dP(\lambda)\chi_{f^{-1}(\lambda_0)} = \lambda_0 \chi_{f^{-1}(\lambda_0)}$

They also often regard (19) as the expression of a given $\psi$ in the “basis” $\{\chi_{f^{-1}(\lambda)}\}_{\lambda \in \text{ran}(f)}$. This comes from realizing $dP(\lambda)\psi = \chi_{f^{-1}(\lambda)}\psi$. The spectral measure is the only formal way to interpret these associations. Formulas (18) and (19) are rigorous and true, while formulas like

$$T\psi_\lambda = \lambda\psi_\lambda, \quad \psi = \int \langle \psi_\lambda, \psi \rangle \psi_\lambda \, d\lambda$$

are just fantasies that come up in trying to understand what (18) and (19) mean or if you don’t really know operator theory, but nevertheless still useful for understanding heuristic derivations and motivations in quantum theory.

**Example.** A typical example of an unbounded operator is a derivative operator, since in general it’s impossible to have a bound for the derivative depending only on a bound for the function itself (a function may jump very suddenly from 0 to 1). So we’d like to consider $\frac{d}{dx}$. Due to the complex conjugation in the $L^2$ inner-product, this alone wouldn’t be symmetric, so we’ll have to include a $-i$ in front (an $i$ would also work but is slightly inconvenient).

Consider $L^2(S^1)$ (the set of complex-valued $L^2$ periodic functions on $[0,2\pi]$) and the operator $-\frac{i}{2\pi} \frac{d}{dx}$ acting at first only on $C^1$ functions in $L^2(S^1)$. Symmetry comes from integration by parts and no boundary terms due to periodicity:

$$\int_0^{2\pi} \overline{\psi} \left( -\frac{i}{2\pi} \psi' \right) = -\frac{i}{2\pi} \int_0^{2\pi} \overline{\psi} \psi' = \frac{i}{2\pi} \int_0^{2\pi} \overline{\psi} \psi' = \int_0^{2\pi} -\frac{i}{2\pi} \overline{\psi} \psi'$$

It turns out this operator extends to a self-adjoint operator $T$ with domain $H^1(S^1)$, the Sobolev space of $L^2$ periodic functions whose first derivative (defined in a weak sense) is also in $L^2$.

An eigenfunction $\psi$ with eigenvalue $\lambda$ must satisfy

$$-\frac{i}{2\pi} \psi' = \lambda \psi \quad \Rightarrow \quad \psi(x) = Ce^{-i\lambda x}$$
But the periodicity condition forces us to have λ integer (and then we switch to denoting it n instead of λ, because who would denote an integer by λ anyway?). So define

\[ \psi_n(x) = \frac{1}{\sqrt{2\pi}} e^{2\pi i n x} \]

The factor in front was chosen to ensure \( L^2 \) norm one. We can also easily see that these functions are orthogonal to each other. And finally every \( L^2(S^1) \) function \( \psi \) is uniquely expressible by a Fourier series (understood with convergence in \( L^2 \) norm)

\[ \psi(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x} \]

with coefficients

\[ a_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} \psi(k) e^{2\pi i n k} \, dk = \langle \psi_n, \psi \rangle \]

Therefore \( \{\psi_n\}_{n \in \mathbb{Z}} \) is a true basis of eigenfunctions of \( T \).

The spectral theorem can be seen to be true by thinking of \( L^2(S^1) \) as \( \ell^2 = \ell^2(\mathbb{Z}) \), the space of square-summable sequences indexed over the integers (which is just \( L^2(\mathbb{Z}) \)), identifying each \( \psi \) with the sequence \( (a_n)_{n \in \mathbb{Z}} \) of its coefficients. Because

\[ T \left( \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x} \right) = \sum_{n \in \mathbb{Z}} n a_n e^{2\pi i n x} \]

the operator corresponding to \( T \) on the level of \( \ell^2 \) is just the multiplication operator \( (a_n) \mapsto (n a_n) \). The spectral projection \( P_{n_0} \) associated to each singleton \( n_0 \in \mathbb{Z} \) is simply the operator that maps a sequence \( (a_n) \) to the sequence whose only nonzero term is \( a_{n_0} \) at position \( n = n_0 \), and the spectral decomposition of \( T \) is a true discrete sum:

\[ T = \sum_{n \in \mathbb{Z}} n P_n \]

**Example.** The operator \( T : \psi(x) \mapsto x \psi(x) \) on \( L^2(\mathbb{R}) \) is unbounded. The spectral theorem for it is vacuous, since it’s already a multiplication operator. It clearly has no eigenvalues, but its spectrum is all of \( \mathbb{R} \). Physicists like to say that the functions \( \delta(\cdot - x_0) \), for \( x_0 \in \mathbb{R} \), are its eigenvalues, since \( T \delta(\cdot - x_0) = x \delta(\cdot - x_0) = x_0 \delta(\cdot - x_0) \). Also notice how they seem to form an orthonormal set for \( L^2(\mathbb{R}) \), if we think of them as functions equal to zero at all but one point:

- \( \int |\delta(x - x_0)|^2 \, dx = 1 \) in some sense.
- For \( x_0 \neq x_1 \), the ill-defined integral \( \int \delta(x - x_0) \delta(x - x_1) \, dx \) looks like zero.
- Every function \( \psi \in L^2(\mathbb{R}) \) can be written in the form \( \psi(x) = \int \psi(k) \delta(x - k) \, dk \), with the coefficients \( \psi(k) = \langle \delta(\cdot - k), \psi \rangle = \int \psi(x) \delta(x - k) \, dx \) (true in some weak sense of the integral).

The problem is they are not even elements of \( L^2 \) to begin with, so this is not a very rigorous notion of an eigenvalue. Examples like this are what prompt people to believe the false theorem on diagonalization of self-adjoint operators, even though what we have here is a “continuous” basis of eigenfunctions, which by the way should at least make us raise an objection about how this contrasts the fact that the underlying Hilbert space \( L^2 \) is separable!

**Example.** Consider \(-i \frac{d}{dx}\), at first defined on \( C^\infty_c(\mathbb{R}) \) (smooth functions of compact support). Symmetry comes from integration by parts and no boundary terms due to compact support:

\[ \int \overline{\psi(x)} (-i \varphi'(x)) \, dx = -i \int \overline{\psi(x)} \varphi'(x) \, dx = i \int \overline{\psi(x)} \varphi(x) \, dx = \int \overline{-i \psi'(x)} \varphi(x) \, dx \]

This operator extends to a self-adjoint operator \( T \) with domain \( H^1(\mathbb{R}) \). Again this is an example of an operator whose spectrum is everything but without any eigenvalues: there are no \( L^2 \) functions \( \psi \neq 0 \) satisfying \(-i \psi' = \lambda \psi \), since the only such functions are \( \psi\lambda(x) = Ce^{-i\lambda x} \not\in L^2(\mathbb{R}) \).

But again a physicist wouldn’t hesitate to call \( \{\psi\lambda\}_{\lambda \in \mathbb{R}} \) an orthonormal basis (actually only orthogonal; we need to “normalize” by \( \frac{1}{\sqrt{2\pi}} \) first). These functions are definitely not normalized or normalizable:
\[ \frac{1}{\sqrt{2\pi}} \int |e^{-i\lambda x}|^2 \, dx = \frac{1}{\sqrt{2\pi}} \int dx = \infty. \]

Orthogonality between \( \psi_\lambda \) and \( \psi_\mu \) for \( \lambda \neq \mu \) also fails horribly: the integral
\[ \frac{1}{2\pi} \int e^{-ikx} e^{-i\mu x} \, dx = \frac{1}{2\pi} \int e^{-i(\mu - \lambda)x} \, dx \]
doesn’t even exist. However, these properties are actually true in the weak sense, that is, they hold, in the same way as in the previous example, not for these basis elements themselves but for their Fourier transforms (regarded as distributions), which turn out to be just the measures \( \delta(-\lambda) \).

The main selling point for labeling these as an orthogonal basis is again the fact that every \( L^2 \) function \( \psi \) can be expressed in the “continuous sum” form \( \psi = \int \langle \psi_k, \psi \rangle \psi_k \, dk \), where the coefficients are
\[ \langle \psi_k, \psi \rangle = \frac{1}{\sqrt{2\pi}} \int e^{-ikx} \psi(x) \, dx = \frac{1}{\sqrt{2\pi}} \int \psi(x) e^{ikx} \, dx. \]

This is just the relationship between the Fourier transform and its inverse. Thus also this part is only half true: it holds only for \( L^1 \cap L^2 \) functions, which are those for which the Fourier transform really is given by the formula \( \mathcal{F}\psi(k) = \int \psi(x) \psi_k(x) \, dx \). But at least there is some truth to it.

The spectral theorem for this operator can be seen to be true due to the existence of the Fourier transform, which is the unitary operator that converts it into a multiplication operator:
\[ \mathcal{F}(-i\psi')(k) = k\mathcal{F}\psi(k) \quad \Longrightarrow \quad \mathcal{F}T\mathcal{F}^{-1}\varphi(k) = k\varphi(k) \]
for all \( \varphi = \mathcal{F}\psi \in L^2 \).

9. References

In the future I’ll give this section proper formatting.

- Zettilli’s *Quantum Mechanics: Concepts and Applications* has good physical background leading to the development of QM. Easy exercises with answers. The math in the book is questionable.
- Teschl’s *Mathematical Methods in Quantum Mechanics* assumes the QM postulates and proves functional analytic theorems rigorously and concisely, very good presentation, but lacking in exercises.
- Reed and Simon’s *Functional Analysis* (4 volumes) contain all the math needed to talk about unbounded operators in volume 1 and Fourier analysis in volume 2. Many exercises, these are used as textbooks often.
- The Wikipedia pages on the history of quantum mechanics and introduction to quantum mechanics are worth reading, even if lacking in mathematical content.
- Van Dommelen’s *Quantum Mechanics for Engineers* (online at https://ww2.eng.famu.fsu.edu/~dommelen/quantum/style_a/contents.html?q=dommelen/quantum/style_a/contents.html) is a very concise but complete and clear presentation of all things quantum mechanical.
- Lieb’s *The stability of matter: from atoms to stars* (available in the Bull. Amer. Soc., volume 22, number 1, 1990) is a great overview paper of the topics that will be included in section 7 in the future.