Basic quantum mechanics

Teo Banica

Department of Mathematics, University of Cergy-Pontoise, F-95000 Cergy-Pontoise, France. teo.banica@gmail.com

2010 *Mathematics Subject Classification.* 81P05

Key words and phrases. Hydrogen atom, Quantum mechanics

Abstract. This is an introduction to classical and modern quantum mechanics, guided by the main object of the theory, which is the hydrogen atom. We first discuss the early atomic theory of Bohr, and the successful work of Heisenberg, Schrödinger and others, in order to prove the claims regarding hydrogen. Then we discuss the fine structure of hydrogen, first with the corrections coming from the electron spin, and then with the higher corrections, coming from quantum electrodynamics. Afterwards, we embark on a discussion regarding more complicated particle systems, featuring bosons, fermions, and their statistics. Finally, we go back to atoms, with a detailed presentation of the periodic table, ions, isotopes, and an introduction to molecules and quantum chemistry.

Preface

Quantum mechanics is a relatively new physics discipline, dating back to the 1920s, that is, just one century ago. Expect of course not to understand much. And also, with the whole theory lacking solid foundations. I'm saying this in view of what happened to other branches of physics, such as classical mechanics, fluid mechanics, optics, electromagnetism, or thermodynamics, whose development took a long long time, and certainly much more than a century, in each case. This is how science goes, we humans are kind of slow, and often unwilling, as individuals, to change our mind, and it always takes quite a few generations of scientists, in order to properly understand something.

Nevermind. You are certainly here for learning quantum mechanics, and this is what we will do, learning quantum mechanics. With the comment, however, that what we will be learning here will be not "the" quantum mechanics, that is, the true quantum mechanics, that of this real world, but rather some sort of blurred version of the theory, which is the one available to us, humans, as of now, beginning of the 21st century.

In practice now, the first question is of course, why quantum mechanics? Is there anything wrong with the traditional branches of physics, that is, classical mechanics, electromagnetism and so on? What are the phenomena that these cannot explain?

In answer, nothing much is wrong, because as long as you look at this world with a relaxed eye, and even under a good microscope, or start doing some basic engineering work, things just fine with traditional physics, you won't need more than that.

However, and here comes the point, while with our usual human lives, and terrestrial environment, and senses, and occupations, we are basically fine with traditional physics, there might be some other interesting things to be known, going beyond this. Is matter made of some sort of atoms? Can we explain chemistry? What about electricity, and magnetism? What exactly is light, and what is its precise relation with matter? What about the relation between light and heat? What about stars? And so on.

Quantum mechanics attempts to answer all these questions, which are all a bit philosophical. And of course, with the idea in mind that all this new knowledge, once worked out, can lead to some applications, that is, to some new, useful technologies too.

The idea of quantum mechanics is very simple: solve the hydrogen atom, and everything else, answers to the above questions, will come naturally, afterwards.

To be more precise, following various theoretical and experimental findings, it became more and more obvious, in the years after 1900, that hydrogen should consist of molecules, themselves made of atoms, and with each atom formed by a negative charge, called electron, spinning around a positive charge, called proton. So, all in all, something which looks quite simple, corresponding to a 2-body problem in electrodynamics.

However, and comes here the surprise, the Maxwell equations for electrodynamics, while certainly true and very useful at our usual scales, are wrong at that small scales, and cannot properly explain the movement of the electron around the proton. And, in order to fix these equations, something terribly complicated must be invented, and then fine-tuned, and fine-tuned again, and so on, to the point that no one really understands anything, and with all this amounting in some sort of big earthquake on all the physics that we knew, since Newton, and providing too answers to the above philosophical questions, and finally, not to forget this, having some applications too, namely nuclear energy and weapons. Welcome to Hell, you would say. Well, welcome to quantum mechanics.

The present book will be an introduction to this. We will first discuss the early atomic theory of Bohr, and the successful work of Heisenberg, Schrödinger and others, in order to prove the claims regarding hydrogen. Then we will discuss the fine structure of hydrogen, first with the corrections coming from the electron spin, and then with the higher corrections, coming from quantum electrodynamics. Afterwards, we will embark on a discussion regarding more complicated particle systems, featuring bosons, fermions, and their statistics. Finally, we will go back to atoms, with a detailed presentation of the periodic table, and an introduction to molecules and quantum chemistry.

Many thanks to the book of Griffiths [43], from where I personally learned all these things. Thanks as well to my cats, although having been in theoretical quantum physics for a while, I only started to learn the fundamentals quite late, with the main aim of understanding the cat intelligence, speed, and appearing and disappearing tricks. And, still working on that, this quantum mechanics seems damn difficult, for us humans.

Cergy, January 2025
Teo Banica

Contents

Part I

Hydrogen atom

There is no going back
I can't stop feeling now
I am not the same
I'm growing up again

CHAPTER 1

Atomic theory

1a. Atomic theory

We first need to talk about light, which will be our main tool for getting into quantum mechanics. And with this being of course a general principle from usual life too. Going down to the cave, or up to the barn? Turn on the light first, to see what's there.

Many things can be said about light, ranging from simple and intuitive to fairly complicated. Philosophically speaking, here is what we need to know:

FACT 1.1 (Einstein principles). *The following happen:*

(1) *Light travels in vacuum at* $c = 299\,792\,458$ m/s.
(2) *This speed* c *is the same for all inertial observers.*
(3) *In non-vacuum, the light speed is lower,* $v < c$.
(4) *Nothing can travel faster than light,* $v \not> c$.

To be more precise here, (1) is something known from long ago, with that precise figure being by definition exact, as per latest the latest SI regulations, defining the meter in terms of c. Regarding (2), this is a tricky thing, due to Einstein, and with the whole theory coming from it, relativity, being verified experimentally. Regarding (3), this is something known for a long time too, and more on it in a moment. As for (4), this is part of Einstein's relativity too, and with the whole thing being verified experimentally.

All the above does not tell us what light is, but we have here:

FACT 1.2 (Maxwell theory). *In regions of space where there is no charge or current present the Maxwell equations for electrodynamics read*

$$< \nabla, E >=< \nabla, B >= 0$$

$$\nabla \times E = -\dot{B} \quad , \quad \nabla \times B = \dot{E}/c^2$$

and both the electric field E *and magnetic field* B *are subject to the wave equation*

$$\ddot{\varphi} = c^2 \Delta \varphi$$

where $\Delta = \sum_i d^2/dx_i^2$ *is the Laplace operator, and* c *is the speed of light.*

To be more precise here, the general Maxwell equations for electrodynamics, which appear as a mixture of math and physics, are globally a physics fact. In the circumstances in the statement these reduce to the equations above, and the verification of the wave equation is a theorem, obtained via some calculus. The speed of light c has something to do with all this via the key formula $\varepsilon_0\mu_0 = 1/c^2$ from the Biot-Savart law, with this numeric observation, namely that we have indeed $\varepsilon_0\mu_0 = 1/c^2$, being due to Maxwell.

So, what is light? Light is the wave predicted by Fact 1.2, with its properties being in tune with what Fact 1.1 says, and with an important extra property being that it depends on a real positive parameter, that can be called, upon taste, frequency, wavelength, or color. And in what regards the creation of light, the mechanism here is as follows:

FACT 1.3. *An accelerating or decelerating charge produces electromagnetic radiation, called light, whose frequency and wavelength can be explicitely computed.*

This phenomenon can be observed is a variety of situations, such as the usual light bulbs, where electrons get decelerated by the filament, acting as a resistor, or in usual fire, which is a chemical reaction, with the electrons moving around, as they do in any chemical reaction, or in more complicated machinery like nuclear plants, particle accelerators, and so on, leading there to all sorts of eerie glows, of various colors.

In view of the above, and especially of the light bulb example, a natural question appears: what about a resistor which is not a light bulb filament, what happens to the light produced there? This is a good question, and we already know a part of answer to it, from the Joule law, saying that the resistor will start heating. The second part of it, that we will discuss in a moment, states that responsible for heat is, guess who, light again, but this time in non-visible wavelenghts, typically IR.

With this discussed, let us go back now to the wave equation $\ddot{\varphi} = v^2\Delta\varphi$ from Fact 1.2, and try to understand its simplest solutions. In 1D, the situation is as follows:

THEOREM 1.4. *The 1D wave equation, with speed v, namely*

$$\ddot{\varphi} = v^2\,\frac{d^2\varphi}{dx^2}$$

has as basic solutions the following functions,

$$\varphi(x) = A\cos(kx - wt + \delta)$$

with A being called amplitude, $kx-wt+\delta$ being called the phase, k being the wave number, w being the angular frequency, and δ being the phase constant. We have

$$\lambda = \frac{2\pi}{k} \quad , \quad T = \frac{2\pi}{kv} \quad , \quad \nu = \frac{1}{T} \quad , \quad w = 2\pi\nu$$

relating the wavelength λ, period T, frequency ν, and angular frequency w. Moreover, any solution of the wave equation appears as a linear combination of such basic solutions.

PROOF. Our first claim is that the function φ in the statement satisfies indeed the wave equation, with speed $v = w/k$. For this purpose, observe that we have:

$$\ddot{\varphi} = -w^2\varphi \quad , \quad \frac{d^2\varphi}{dx^2} = -k^2\varphi$$

Thus, the wave equation is indeed satisfied, with speed $v = w/k$:

$$\ddot{\varphi} = \left(\frac{w}{k}\right)^2 \frac{d^2\varphi}{dx^2} = v^2\,\frac{d^2\varphi}{dx^2}$$

Regarding now the other things in the statement, all this is basically terminology, which is very natural, when thinking how $\varphi(x) = A\cos(kx - wt + \delta)$ propagates. Finally, the last assertion is something standard, coming from Fourier analysis. □

As a first observation, the above result invites the use of complex numbers. Indeed, we can write the solutions that we found in a more convenient way, as follows:

$$\varphi(x) = Re\left[A\,e^{i(kx-wt+\delta)}\right]$$

And we can in fact do even better, by absorbing the quantity $e^{i\delta}$ into the amplitude A, which becomes now a complex number, and writing our formula as:

$$\varphi = Re(\widetilde{\varphi}) \quad , \quad \widetilde{\varphi} = \widetilde{A}e^{i(kx-wt)}$$

Moving ahead now towards electromagnetism and 3D, let us formulate:

DEFINITION 1.5. *A monochromatic plane wave is a solution of the 3D wave equation which moves in only 1 direction, making it in practice a solution of the 1D wave equation, and which is of the special from found in Theorem 1.4, with no frequencies mixed.*

In other words, we are making here two assumptions on our wave. First is the 1-dimensionality assumption, which gets us into the framework of Theorem 1.4. And second is the assumption, in connection with the Fourier decomposition result from the end of Theorem 1.4, that our solution is of "pure" type, meaning a wave having a well-defined wavelenght and frequency, instead of being a "packet" of such pure waves.

All this is still mathematics, and making now the connection with physics and electromagnetism, and more specifically with Fact 1.2 and Fact 1.3, we have:

FACT 1.6. *Physically speaking, a monochromatic plane wave is the electromagnetic radiation appearing as in Fact 1.2 and Fact 1.3, via equations of type*

$$E = Re(\widetilde{E}) \quad : \quad \widetilde{E} = \widetilde{E}_0\,e^{i(<k,x>-wt)}$$

$$B = Re(\widetilde{B}) \quad : \quad \widetilde{B} = \widetilde{B}_0\,e^{i(<k,x>-wt)}$$

with the wave number being now a vector, $k \in \mathbb{R}^3$. Moreover, it is possible to add to this an extra parameter, accounting for the possible polarization of the wave.

To be more precise, what we are doing here is to import the conclusions of our mathematical discussion so far, from Theorem 1.4 and Definition 1.5, into the context of our original physics discussion, from Fact 1.3 and Fact 1.4. And also to add an extra twist coming from physics, and more specifically from the notion of polarization.

In any case, we have now a decent intuition about what light is, and more on this later, and let us discuss now the examples. The idea is that we have various types of light, depending on frequency and wavelength. These are normally referred to as "electromagnetic waves", but for keeping things simple and luminous, we will keep using the familiar term "light". The classification, in a rough form, is as follows:

Frequency	Type	Wavelength
	–	
$10^{18} - 10^{20}$	γ rays	$10^{-12} - 10^{-10}$
$10^{16} - 10^{18}$	X – rays	$10^{-10} - 10^{-8}$
$10^{15} - 10^{16}$	UV	$10^{-8} - 10^{-7}$
	–	
$10^{14} - 10^{15}$	blue	$10^{-7} - 10^{-6}$
$10^{14} - 10^{15}$	yellow	$10^{-7} - 10^{-6}$
$10^{14} - 10^{15}$	red	$10^{-7} - 10^{-6}$
	–	
$10^{11} - 10^{14}$	IR	$10^{-6} - 10^{-3}$
$10^{9} - 10^{11}$	microwave	$10^{-3} - 10^{-1}$
$1 - 10^{9}$	radio	$10^{-1} - 10^{8}$

Observe the tiny space occupied by the visible light, all colors there, and the many more missing, being squeezed under the $10^{14} - 10^{15}$ frequency banner. Here is a zoom on that part, with of course the remark that all this, colors, is something subjective:

Frequency THz $= 10^{12}$ Hz	Color	Wavelength nm $= 10^{-9}$ m
	–	
$670 - 790$	violet	$380 - 450$
$620 - 670$	blue	$450 - 485$
$600 - 620$	cyan	$485 - 500$
$530 - 600$	green	$500 - 565$
$510 - 530$	yellow	$565 - 590$
$480 - 510$	orange	$590 - 625$
$400 - 480$	red	$625 - 750$

Outside visible light we have, as you probably know it, UV on higher frequencies, and IR on lower frequencies. At the high frequency end we have X-rays, that you surely know about too, and γ rays, which are usually associated with various bad things, such as thunderstorms, solar flares, and small bugs with our nuclear energy technology.

As for the lower frequency end of the scale, first we have microwaves, but if you love physics and chemistry you should learn some cooking, that's first-class chemistry, that you can practice every day. And then we have all sorts of radio wavelenghts, including FM, followed by AM, and then by several more obscure low-frequency waves.

Importantly, both ends of the table are a bit loose. At the high frequency end there are some restrictions coming from quantum mechanics, and more on them later. As for the low frequency end, what's wave and what's not is a bit of a philosophical question, but which is actually not that philosophical, because waves having huge wavelengths can easily turn around mountains, full countries and so on, and so are of military interest. Secret research here, more of engineering type of course, is still ongoing.

Back now to our business, with all the above in hand, we can do some optics. Light usually comes in "bundles", with waves of several wavelenghts coming at the same time, from the same source, and the first challenge is that of separating these wavelenghts.

In order to discuss this, let us start with the following fact:

FACT 1.7. *Inside a linear, homogeneous medium, where there is no free charge or current present, the Maxwell equations for electrodynamics read*

$$< \nabla, E >=< \nabla, B >= 0$$

$$\nabla \times E = -\dot{B} \quad , \quad \nabla \times B = \varepsilon\mu\dot{E}$$

with E, B being as before the electric and the magnetic field, and with $\varepsilon > \varepsilon_0$ and $\mu > \mu_0$ being the electric permittivity and magnetic permeability of the medium.

Observe that this is precisely the first part of Fact 1.2, with the vacuum constants ε_0, μ_0 being replaced by their versions ε, μ, concerning the medium in question. In what regards now the second part of Fact 1.2, which was a theorem, we have:

THEOREM 1.8. *Inside a linear, homogeneous medium, where there is no free charge or free current present, both E and B are subject to the wave equation*

$$\ddot{\varphi} = v^2\Delta\varphi$$

with v being the speed of light inside the medium, given by

$$v = \frac{c}{n} \quad : \quad n = \sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}}$$

with the quantity on the right $n > 1$ being called refraction index of the medium.

PROOF. This is something that we know well in vacuum, and the proof in general is identical, with the resulting speed being $v = 1/\sqrt{\varepsilon\mu}$, which leads to the above. □

As a first observation here, while the above is something quite trivial, mathematically speaking, from the physical viewpoint we are here into complicated things. Materials can be transparent or opaque, with the distinction between them being something very subtle, and advanced, and Theorem 1.8 obviously deals with the transparent case.

In short, we are here inside advanced materials theory, that we cannot really understand, with our knowledge so far. In what follows we will be interested in transparent materials only, such as glass. Regarding the other materials, such as rock, let us just mention that light dissapears inside them, converted into heat. Of course glass heats too when light crosses it, with this being related to $v < c$ inside it. More on this later.

Next in line, and for interest for us, we have:

FACT 1.9. *When traveling through a material, and hitting a new material, some of the light gets reflected, at the same angle, and some of it gets refracted, at a different angle, depending both on the old and the new material, and on the wavelength.*

Again, this is something deep, and very old as well, and there are many things that can be said here, ranging from various computations based on the Maxwell equations, to all sorts of considerations belonging to advanced materials theory.

As a basic formula here, we have the famous Snell law, which relates the incidence angle θ_1 to the refraction angle θ_2, via the following simple formula:

$$\frac{\sin\theta_2}{\sin\theta_1} = \frac{n_1(\lambda)}{n_2(\lambda)}$$

Here $n_i(\lambda)$ are the refraction indices of the two materials, adjusted for the wavelength, and with this adjustment for wavelength being the whole point, which is something quite complicated. For an introduction to all this, we refer for instance to Griffiths [42].

As a simple consequence of the above, we have:

THEOREM 1.10. *Light can be decomposed, by using a prism.*

PROOF. This follows from Fact 1.9. Indeed, when hitting a piece of glass, provided that the hitting angle is not 90°, the light will decompose over the wavelenghts present, with the corresponding refraction angles depending on these wavelengths. And we can capture these split components at the exit from the piece of glass, again deviated a bit, provided that the exit surface is not parallel to the entry surface. And the simplest device doing the job, that is, having two non-parallel faces, is a prism. □

Good news, with this in hand, we can now talk about spectroscopy:

FACT 1.11. *We can study events via spectroscopy, by capturing the light the event has produced, decomposing it with a prism, carefully recording its "spectral signature", consisting of the wavelenghts present, and their density, and then doing some reverse engineering, consisting in reconstructing the event out of its spectral signature.*

This is the main principle of spectroscopy, and applications, of all kinds, abound. In practice, the mathematical tool needed for doing the "reverse engineering" mentioned above is the Fourier transform, which allows the decomposition of packets of waves, into monochromatic components. Finally, let us mention too that, needless to say, the event can be reconstructed only partially out of its spectral signature.

As a conclusion to all this, we have learned many things about light, and in particular the method of spectroscopy. And so, we can now go back to atoms, and study them, simply by heating them, and then observing the color of the light that they produce.

So, getting now back to atoms, there is a long story here, involving many discoveries of many people, around 1890-1900, focusing on hydrogen H. We will present here things a bit retrospectively, as to bet fit with science as we know it now, and with the present book. First on our list is the following discovery, by Lyman in 1906:

FACT 1.12 (Lyman). *The hydrogen atom has spectral lines given by the formula*

$$\frac{1}{\lambda} = R\left(1 - \frac{1}{n^2}\right)$$

where $R \simeq 1.097 \times 10^7$ *and* $n \geq 2$*, which are as follows,*

n	Name	Wavelength	Color
	–	–	
2	α	121.567	UV
3	β	102.572	UV
4	γ	97.254	UV
$\vdots$	$\vdots$	$\vdots$	$\vdots$
∞	limit	91.175	UV

called Lyman series of the hydrogen atom.

Observe that all the Lyman series lies in UV. Due to this fact, namely the invisibility of UV to the human eye, this series, while theoretically being the most important, for certain reasons to be explained later, was discovered only second.

The first discovery, which was the big one, and the breakthrough, was by Balmer, the founding father of all this, back in 1885, in the visible range, as follows:

FACT 1.13 (Balmer). *The hydrogen atom has spectral lines given by the formula*

$$\frac{1}{\lambda} = R\left(\frac{1}{4} - \frac{1}{n^2}\right)$$

where $R \simeq 1.097 \times 10^7$ *and* $n \geq 3$*, which are as follows,*

n	Name	Wavelength	Color
	–	–	
3	α	656.279	red
4	β	486.135	aqua
5	γ	434.047	blue
6	δ	410.173	violet
7	ε	397.007	UV
$\vdots$	$\vdots$	$\vdots$	$\vdots$
∞	limit	346.600	UV

called Balmer series of the hydrogen atom.

So, this was Balmer's original result, which started everything, and with his original wavelength formula being in fact something equivalent to the above formula, but a bit more complicated, as follows, with $B \simeq 3.645 \times 10^{-7}$ being the Balmer constant:

$$\lambda = \frac{Bn^2}{n^2 - 4}$$

As a third main result now, this time in IR, due to Paschen in 1908, we have:

FACT 1.14 (Paschen). *The hydrogen atom has spectral lines given by the formula*

$$\frac{1}{\lambda} = R\left(\frac{1}{9} - \frac{1}{n^2}\right)$$

where $R \simeq 1.097 \times 10^7$ *and* $n \geq 4$*, which are as follows,*

n	Name	Wavelength	Color
	–	–	
4	α	1875	IR
5	β	1282	IR
6	γ	1094	IR
$\vdots$	$\vdots$	$\vdots$	$\vdots$
∞	limit	820.4	IR

called Paschen series of the hydrogen atom.

Observe the striking similarity between the above three results. In fact, we have here the following fundamental, grand result, due to Rydberg in 1888, based on the Balmer series, and with later contributions by Ritz in 1908, using the Lyman series as well:

CONCLUSION 1.15 (Rydberg, Ritz). *The spectral lines of the hydrogen atom are given by the Rydberg formula, depending on integer parameters $n_1 < n_2$,*

$$\frac{1}{\lambda_{n_1n_2}} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

with R being the Rydberg constant for hydrogen, which is as follows:

$$R \simeq 1.096\ 775\ 83 \times 10^7$$

These spectral lines combine according to the Ritz-Rydberg principle, as follows:

$$\frac{1}{\lambda_{n_1n_2}} + \frac{1}{\lambda_{n_2n_3}} = \frac{1}{\lambda_{n_1n_3}}$$

Similar formulae hold for other atoms, with suitable fine-tunings of R.

Here the first part, the Rydberg formula, generalizes the results of Lyman, Balmer, Paschen, which appear at $n_1 = 1, 2, 3$, at least retrospectively. The Rydberg formula predicts further spectral lines, appearing at $n_1 = 4, 5, 6, \ldots$, and these were discovered later, by Brackett in 1922, Pfund in 1924, Humphreys in 1953, and others aftwerwards, with all these extra lines being in far IR. The simplified complete table is as follows:

n_1	n_2	Series name	Wavelength $n_2 = \infty$	Color $n_2 = \infty$
		–	–	
1	$2 - \infty$	Lyman	91.13 nm	UV
2	$3 - \infty$	Balmer	364.51 nm	UV
3	$4 - \infty$	Paschen	820.14 nm	IR
		–	–	
4	$5 - \infty$	Brackett	1458.03 nm	far IR
5	$6 - \infty$	Pfund	2278.17 nm	far IR
6	$7 - \infty$	Humphreys	3280.56 nm	far IR
...	...	...	...	...

Regarding the last assertion, concerning other elements, this is something conjectured and partly verified by Ritz, and fully verified and clarified later, via many experiments, the fine-tuning of R being basically $R \to RZ^2$, where Z is the atomic number.

But from a theoretical physics viewpoint, the main result remains the middle assertion, called Ritz-Rydberg combination principle. This is something at the same time extremely simple, and completely puzzling, the informal conclusion being as follows:

THOUGHT 1.16. *The simplest observables of the hydrogen atom, combining via*

$$\frac{1}{\lambda_{n_1n_2}} + \frac{1}{\lambda_{n_2n_3}} = \frac{1}{\lambda_{n_1n_3}}$$

look like quite weird quantities. Why wouldn't they just sum normally.

Getting now to quantum mechanics, and to our dreams about it, formulated before, well, good news, we have some serious data here. These spectral lines are basic and beautiful, obviously of quantized type, and in order to get started with our theory, we first need to solve the puzzle of the Ritz-Rydberg combination principle.

But, how to do this? Fortunately, matrix theory comes to the rescue, as follows:

THOUGHT 1.17. *The Ritz-Rydberg combination principle reminds the formula*

$$e_{n_1n_2}e_{n_2n_3} = e_{n_1n_3}$$

for the usual matrix units, which are the elementary matrices given by

$$e_{ij} : e_j \to e_i$$

perhaps taken in infinite dimensions, as to allow infinite-ranging indices.

This looks certainly interesting, and actually reminds a well-known speculation in relation with relativity. Remember the Galileo addition formula for relative speeds, which gets converted into the Einstein addition formula for the same relative speeds? Well, the reasoning there was as follows, involving the same sort of weird additions as above:

$$v_{AB} +_g v_{BC} = v_{AC} \quad : \quad v_{AC} = v_{AB} + v_{BC}$$

$$\implies \quad v_{AB} +_e v_{BC} = v_{AC} \quad : \quad v_{AC} = \frac{v_{AB} + v_{BC}}{1 + v_{AB}v_{BC}/c^2}$$

In short, be that with matrix theory, or with relative speeds in classical mechanics or relativity, we are in familiar territory here, and we can start dreaming of:

THOUGHT 1.18. *Observables in quantum mechanics should be some sort of infinite matrices, generalizing the Lyman, Balmer, Paschen lines of the hydrogen atom, and multiplying between them as the matrices do, as to produce further observables.*

In practice now, as a main problem that we would like to solve, we have the understanding the intimate structure of matter, at the atomic level.

There is of course a long story here, regarding the intimate structure of matter, going back centuries and even millennia ago, and our presentation here will be quite simplified. As a starting point, since we need a starting point, to start with, let us agree on:

CLAIM 1.19. *Ordinary matter is made of small particles called atoms, with each atom appearing as a mix of even smaller particles, namely protons* $+$*, neutrons* 0 *and electrons* $-$*, with the same number of protons* $+$ *and electrons* $-$*.*

As a first observation, this is something which does not look obvious at all, with probably lots of work, by many people, being involved, as to lead to this claim. And so it is. The story goes back to the discovery of charges and electricity, which were attributed

to a small particle, the electron $-$. Now since matter is by default neutral, this naturally leads to the consideration to the proton $+$, having the same charge as the electron.

But why should be these electrons $-$ and protons $+$ that small? And also, what about the neutron 0? These are not easy questions, and the fact that indeed it is so came from several clever experiments, due to Thomson, Rutherford and others. To be more precise, bombarding matter with other pieces of matter, accelerated to the extremes, or submitting it to huge electric and magnetic fields, does work, and leads to the conclusion that these small beasts $+, 0, -$ exist indeed, in agreement with Claim 1.19.

Of particular importance here was as well the radioactivity theory of Becquerel and Pierre and Marie Curie, involving this time such small beasts, or perhaps some related radiation, peeling off by themselves, in heavy elements such as uranium ${}_{92}\mathrm{U}$, polonium ${}_{84}\mathrm{Po}$ and radium ${}_{88}\mathrm{Ra}$. And there was also Einstein's work on the photoelectric effect, light interacting with matter, suggesting that even light itself might have associated to it some kind of particle, called photon. All this goes of course beyond Claim 1.19, with further particles involved, and more on this later, but as a general idea, all this deluge of small particle findings, all coming around 1900-1910, further solidified Claim 1.19.

So, taking now Claim 1.19 for granted, how are then the atoms organized, as mixtures of protons $+$, neutrons 0 and electrons $-$? The answer here lies again in the above-mentioned experiments of Thomson, Rutherford and others, which led to:

Claim 1.20. *The atoms are formed by a core of protons $+$ and neutrons 0, surrounded by a cloud of electrons $-$, gravitating around the core.*

This is a considerable advance, because we are now into familiar territory, namely some kind of mechanics. Remember from classical mechanics the planets orbiting around the Sun, on ellipses, as found by Kepler, and proved by Newton? Well, the same should happen with electrons orbiting around the core, but this time due to the Coulomb force. And with this in mind, all the pieces of our puzzle start fitting together, and lead to:

Claim 1.21 (Bohr and others). *The atoms are formed by a core of protons and neutrons, surrounded by a cloud of electrons, basically obeying to a modified version of electromagnetism. And with a fine mechanism involved, as follows:*

(1) *The electrons are free to move only on certain specified elliptic orbits, labelled $1, 2, 3, \ldots$, situated at certain specific heights.*
(2) *The electrons can jump or fall between orbits $n_1 < n_2$, absorbing or emitting light and heat, that is, electromagnetic waves, as accelerating charges.*
(3) *The energy of such a wave, coming from $n_1 \to n_2$ or $n_2 \to n_1$, is given, via the Planck viewpoint, by the Rydberg formula, applied with $n_1 < n_2$.*
(4) *The simplest such jumps are those observed by Lyman, Balmer, Paschen. And multiple jumps explain the Ritz-Rydberg formula.*

And isn't this beautiful. Moreover, some further claims, also by Bohr and others, are that the theory can be further extended and fine-tuned as to explain many other phenomena, such as the above-mentioned findings of Einstein, and of Becquerel and Pierre and Marie Curie, and generally speaking, all the physics and chemistry known.

And the story is not over here. Following now Heisenberg, the next claim is that the underlying math in all the above can lead to a beautiful axiomatization of quantum mechanics, as a "matrix mechanics", along the lines of Thought 1.17.

1b.

1c.

1d.

1e. Exercises

Exercises:

Exercise 1.22.

Exercise 1.23.

Exercise 1.24.

Exercise 1.25.

Exercise 1.26.

Exercise 1.27.

Exercise 1.28.

Exercise 1.29.

Bonus exercise.

CHAPTER 2

Schrödinger equation

2a. Schrödinger equation

Getting started now, following Schrödinger, let us forget about exact things, and try to investigate the hydrogen atom statistically. We have here the following question:

QUESTION 2.1. *In the context of the hydrogen atom, assuming that the proton is fixed, what is the probability density $\varphi_t(x)$ of the position of the electron e, at time t,*

$$P_t(e \in V) = \int_V \varphi_t(x)dx$$

as function of an intial probability density $\varphi_0(x)$? Moreover, can the corresponding equation be solved, and will this prove the Bohr claims for hydrogen, statistically?

In order to get familiar with this question, let us first look at examples coming from classical mechanics. In the context of linear motion, with speed v, we have:

$$\varphi_t(x) = \varphi_0(x) + vt$$

More generally, assuming that we have a particle whose position at time t is given by $x_0 + \gamma(t)$, the evolution of the probability density will be given by:

$$\varphi_t(x) = \varphi_0(x) + \gamma(t)$$

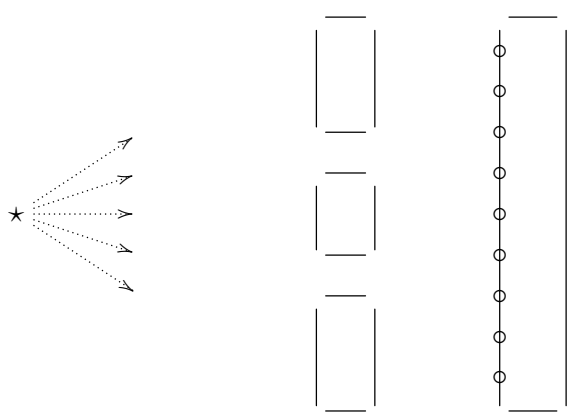
These examples are somewhat trivial, of course not in relation with the computation of γ, usually a difficult question, but in relation with our questions, and do not apply to the electron. The point indeed is that, in what regards the electron, we have:

FACT 2.2. *In respect with various simple interference experiments:*

(1) *The electron is definitely not a particle in the usual sense.*
(2) *But in most situations it behaves exactly like a wave.*
(3) *But in other situations it behaves like a particle.*

So, here we go again with puzzles. These experiments are nicely described, with extensive comments, in Feynman's book [35]. In what follows, we will present them quickly. We will just need, for our purposes here, the first 4 experiments in the series,

which are the most important. These are performed with a machinery as follows:



To be more precise, on the left we have a multi-purpose gun $\star$, which can shoot bullets, water waves, or electrons. On the middle we have a wall with two holes in it. On the right we have a solid wall, with sensors $\circ$, adapted to the matter that we are shooting.

The first experiment, performed with 22 cal ammo, assumed to be idealized, as to be indestructible, and we refer again to Feynman [35] for full details, goes as follows:

EXPERIMENT 2.3. *When shooting bullets, the density function on the right wall, stopping them, as recorded by our sensors there, is given by*

$$\varphi = \varphi_1 + \varphi_2$$

where φ_1 is the same density, but measured with the lower hole closed, and φ_2 is also the same density, but measured with the upper hole closed.

Nothing surprising so far. Now let us shoot water waves, or rather assume that our gun $\star$ is a wave source. For this experiment, the sensors at right are set to measure the energy of the incoming wave, which is proportional to the square of the height.

The experiment, best performed with our favorite drinkable, tap water, gives:

EXPERIMENT 2.4. *When shooting waves, the energy density functions at right, measured with one or the other hole open, or both holes open, are related by*

$$\varphi_1 = |\psi_1|^2 \quad , \quad \varphi_2 = |\psi_1|^2 \quad , \quad \varphi = |\psi_1 + \psi_2|^2$$

where $\psi_1, \psi_2 \in \mathbb{C}$ are the amplitudes of the waves passing through one hole, with the other hole closed. This phenomenon and formula of φ are due to interference.

This is again, something not surprising, that we know, coming from the fact that two colliding waves can add up in various ways, depending on their phases.

Now let us shoot electrons. At first sight, these behave like particles, because our sensors beep for them one at the time. However, when examining the results regarding probability distributions, these don't add up as for bullets, the conclusions being:

EXPERIMENT 2.5. *When shooting electrons, these come up one at the time, exactly as bullets. However, in what regards the density functions, these don't add up:*

$$\varphi \neq \varphi_1 + \varphi_2$$

Thus, we have some interference, and most likely the correct formula is

$$\varphi_1 = |\psi_1|^2 \quad , \quad \varphi_2 = |\psi_1|^2 \quad , \quad \varphi = |\psi_1 + \psi_2|^2$$

with $\psi_1, \psi_2 \in \mathbb{C}$ being certain amplitudes, exactly as for the waves.

This is a bit surprising, showing that the electrons have a mix of particle and wave behavior, at least with respect to this experiment. Let us also mention too that, contrary to the previous two experiments which are simple and real, this is rather a Gedankenexperiment, and so the wave formulae are to be taken with care. See Feynman [35].

Finally, as a last experiment, again with electrons, we have:

EXPERIMENT 2.6. *When shooting electrons as before, but by putting a light bulb behind one hole, whose light is scattered by electrons passing through that hole:*

$$\varphi = \varphi_1 + \varphi_2$$

That is, observing the electrons passing through one hole, via them scattering light, has killed the interference process, and we have now usual particles, like bullets.

And this is probably the most surprising experiment of them all. Indeed, the fact that in Experiment 2.5 we have particles when counting and waves when looking at densities might seem odd, but after all, why not. So these are our beasts, electrons, and this is how their properties are, a bit odd, but at least we know one thing. However, what happens now seems to defy any logic. Observing the electrons has changed their properties, and that's how things are. Welcome to quantum mechanics.

Getting back now to the Schrödinger question, all this suggests to use, as for the waves, an amplitude function $\psi_t(x) \in \mathbb{C}$, related to the density $\varphi_t(x) > 0$ by the formula $\varphi_t(x) = |\psi_t(x)|^2$. So, let us reformulate Question 2.1, in the following way:

QUESTION 2.7. *In the context of the hydrogen atom, assuming that the proton is fixed, what is the amplitude function $\psi_t(x)$ of the position of the electron e, at time t,*

$$P_t(e \in V) = \int_V |\psi_t(x)|^2 dx$$

as function of an intial amplitude function $\psi_0(x)$? Moreover, can the corresponding equation be solved, and will this prove the Bohr claims for hydrogen, statistically?

Mathematically, what we did here is to replace the density $\varphi_t(x) > 0$ by the amplitude function $\psi_t(x) \in \mathbb{C}$. Not that a big deal, you would say, because the two are related by simple formulae as follows, with $\theta_t(x)$ being an arbitrary phase function:

$$\varphi_t(x) = |\psi_t(x)|^2 \quad , \quad \psi_t(x) = e^{i\theta_t(x)}\sqrt{\varphi_t(x)}$$

However, experience with math shows that such manipulations can be crucial, raising for instance the possibility that the amplitude function satisfies some simple equation, while the density itself, maybe not. So, let us hope for this to happen.

And this is what happens indeed. Schrödinger was led in this way to:

CLAIM 2.8 (Schrödinger). *In the context of the hydrogen atom, the amplitude function of the electron $\psi = \psi_t(x)$ is subject to the Schrödinger equation*

$$ih\dot{\psi} = -\frac{h^2}{2m}\Delta\psi + V\psi$$

m being the mass, h the modified Planck constant, and V the Coulomb potential of the proton. The same holds for movements of the electron under an arbitrary potential V.

Observe the similarity with the wave equation $\ddot{\varphi} = v^2\Delta\varphi$, and with the heat equation $\dot{\varphi} = \alpha\Delta\varphi$ too. There might be of course some speculations to be made here, but passed that, this is certainly not your easy to decipher equation. So, where does this equation come from? Is there a way of deducing it from simpler principles? And so on.

Generally speaking, however, any axiomatic explanation for the Schrödinger equation can only introduce some possible mistakes in our theory. And so we are led by precaution to the following preliminary answer, to such questions, that you might have:

COMMENT 2.9. *The Schrödinger equation comes from Schrödinger.*

And please do not take this as a joke. We are mainly interested in solving the hydrogen atom, and the Schrödinger equation can only solve it, via some calculus. So why not enjoying this, solving the hydrogen atom by using this equation, and see later what further things, beyond Schrödinger, can be said about quantum mechanics.

This being said, before getting into computations, let us discuss however, a bit in advance, some possible ways of getting into the Schrödinger equation. We first have:

COMMENT 2.10. *The Schrödinger equation appears naturally from an abstract claim of de Broglie, regarding the precise wave properties of the electron.*

To be more precise here, the above-mentioned abstract claim of de Broglie leads to the following equations for the wave function of a free electron:

$$\psi_t = e^{-iEt/h}\psi_0 \quad , \quad E\psi_0 = -\frac{h^2}{2m}\Delta\psi_0$$

Now in the context of movement under a time-independent potential V, as is the potential coming from the proton, these equations can be naturally modified into:

$$\psi_t = e^{-iEt/h}\psi_0 \quad , \quad E\psi_0 = -\frac{h^2}{2m}\Delta\psi_0 + V\psi_0$$

But this is exactly the simplified form of the general Schrödinger equation from Claim 2.8, in the case of a time-independent potential, as we will soon see.

We have as well a second method for getting into the Schrödinger equation, a bit more powerful, but based on more powerful assumptions too, as follows:

COMMENT 2.11. *The Schrödinger equation appears naturally by invoking a bit of matrix mechanics of Heisenberg type, and the Hamiltonian.*

To be more precise here, according to the viewpoint of Heisenberg, the total energy, or Hamiltonian, $H = T + V$ is represented by the following "operator":

$$\widehat{H} = -\frac{h^2\Delta}{2m} + V$$

And in terms of this operator, the Schrödinger equation simply appears as:

$$ih\dot{\psi} = \widehat{H}\psi$$

This is actually the explanation offered by Schrödinger himself in his paper, and we will comment on this a bit later, when having a better knowledge of the subject. We refer also to Feynman [35], Griffiths [43], Weinberg [93] for more on all this.

Now, let us go back to the Schrödinger equation from Claim 2.8, and try to solve it. Let us start with some computations. As a first question, we would like to see how the probability density $\varphi = |\psi|^2$ evolves in time, and we have here:

PROPOSITION 2.12. *In the context of the general Schrödinger equation,*

$$ih\dot{\psi} = -\frac{h^2}{2m}\Delta\psi + V\psi$$

we have the following formula,

$$\dot{\varphi} = \frac{ih}{2m}\left(\Delta\psi\cdot\bar{\psi} - \Delta\bar{\psi}\cdot\psi\right)$$

for the time derivative of the probability density function $\varphi = |\psi|^2$.

PROOF. According to the Leibnitz product rule for the derivatives, we have the following formula, for the time derivative of the probability density function:

$$\begin{aligned}\dot{\varphi} &= \frac{d}{dt}|\psi|^2\\ &= \frac{d}{dt}(\psi\bar{\psi})\\ &= \dot{\psi}\bar{\psi}+\psi\dot{\bar{\psi}}\end{aligned}$$

On the other hand, the Schrödinger equation and its conjugate read:

$$\dot{\psi}=\frac{ih}{2m}\left(\Delta\psi-\frac{2m}{h^2}V\psi\right)$$

$$\dot{\bar{\psi}}=-\frac{ih}{2m}\left(\Delta\bar{\psi}-\frac{2m}{h^2}V\bar{\psi}\right)$$

By plugging this data, we obtain the following formula:

$$\dot{\varphi}=\frac{ih}{2m}\left[\left(\Delta\psi-\frac{2m}{h^2}V\psi\right)\bar{\psi}-\left(\Delta\bar{\psi}-\frac{2m}{h^2}V\bar{\psi}\right)\psi\right]$$

But this gives, after simplifying, the following formula:

$$\dot{\varphi}=\frac{ih}{2m}\left(\Delta\psi\cdot\bar{\psi}-\Delta\bar{\psi}\cdot\psi\right)$$

Thus, we are led to the conclusion in the statement. □

As an important application now of Proposition 2.12, which is of main theoretical interest, we have the following key result:

THEOREM 2.13. *The general Schrödinger equation, namely*

$$ih\dot{\psi}=-\frac{h^2}{2m}\Delta\psi+V\psi$$

conserves probability amplitudes, in the sense that we have

$$\int_{\mathbb{R}^3}|\psi_0|^2=1\implies\int_{\mathbb{R}^3}|\psi_t|^2=1$$

in agreement with the basic probabilistic requirement, $P=1$ overall.

PROOF. According to the formula in Proposition 2.12, we have the following computation, for the time derivative of the quantity that we are interested in:

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^3}|\psi|^2\,dx &= \int_{\mathbb{R}^3}\frac{d}{dt}|\psi|^2\,dx\\ &= \int_{\mathbb{R}^3}\dot{\varphi}\,dx\\ &= \frac{ih}{2m}\int_{\mathbb{R}^3}\left(\Delta\psi\cdot\bar{\psi}-\Delta\bar{\psi}\cdot\psi\right)dx\end{aligned}$$

Now by remembering the definition of the Laplace operator, we have:

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^3}|\psi|^2\,dx &= \frac{ih}{2m}\int_{\mathbb{R}^3}\sum_i\left(\frac{d^2\psi}{dx_i^2}\cdot\bar{\psi}-\frac{d^2\bar{\psi}}{dx_i^2}\cdot\psi\right)dx\\ &= \frac{ih}{2m}\sum_i\int_{\mathbb{R}^3}\frac{d}{dx_i}\left(\frac{d\psi}{dx_i}\cdot\bar{\psi}-\frac{d\bar{\psi}}{dx_i}\cdot\psi\right)dx\\ &= \frac{ih}{2m}\sum_i\int_{\mathbb{R}^2}\left[\frac{d\psi}{dx}\cdot\bar{\psi}-\frac{d\bar{\psi}}{dx}\cdot\psi\right]_{-\infty}^{\infty}\frac{dx}{dx_i}\\ &= \frac{ih}{2m}\sum_i\int_{\mathbb{R}^2}0\,\frac{dx}{dx_i}\\ &= 0\end{aligned}$$

Here we have used at the end the assumption, which is physically speaking, something reasonable, that the wave function and its derivatives vanish at ∞. Now with this in hand, since the quantity under consideration is constant, we obtain the result. □

Let us do now some computations, in order to get some insight into the quantum mechanics of the particle, as dictated by the Schrödinger equation. We first have:

THEOREM 2.14. *The average position and momentum of the particle are*

$$<x>=\int_{\mathbb{R}^3}x|\psi|^2\,dx$$

$$<p>=-ih\int_{\mathbb{R}^3}\nabla\psi\cdot\bar{\psi}\,dx$$

with the convention that the average speed is the derivative of the average position.

PROOF. This follows again by doing some math, as follows:

(1) The formula for the average position $< x >$ is clear from definitions. Regarding now the average speed $< v >$, we have here the following computation:

$$\begin{aligned} < v > &= \frac{d < x >}{dt} \\ &= \int_{\mathbb{R}^3} x \cdot \frac{d}{dt}|\psi|^2 \, dx \\ &= \int_{\mathbb{R}^3} x \, \dot{\varphi} \, dx \\ &= \frac{ih}{2m} \int_{\mathbb{R}^3} x \left(\Delta\psi \cdot \bar{\psi} - \Delta\bar{\psi} \cdot \psi \right) dx \end{aligned}$$

(2) But each of the components can be computed as follows, by taking into account the vanishing formula found in the proof of Theorem 2.13:

$$\begin{aligned} < v >_i &= \frac{ih}{2m} \int_{\mathbb{R}^3} x_i \left(\Delta\psi \cdot \bar{\psi} - \Delta\bar{\psi} \cdot \psi \right) dx \\ &= \frac{ih}{2m} \int_{\mathbb{R}^3} x_i \sum_j \left(\frac{d^2\psi}{dx_j^2} \cdot \bar{\psi} - \frac{d^2\bar{\psi}}{dx_j^2} \cdot \psi \right) dx \\ &= \frac{ih}{2m} \sum_j \int_{\mathbb{R}^3} x_i \left(\frac{d^2\psi}{dx_j^2} \cdot \bar{\psi} - \frac{d^2\bar{\psi}}{dx_j^2} \cdot \psi \right) dx \\ &= \frac{ih}{2m} \int_{\mathbb{R}^3} x_i \left(\frac{d^2\psi}{dx_i^2} \cdot \bar{\psi} - \frac{d^2\bar{\psi}}{dx_i^2} \cdot \psi \right) dx \end{aligned}$$

(3) We can now finish the computation by doing two partial integrations, as follows:

$$\begin{aligned} < v >_i &= \frac{ih}{2m} \int_{\mathbb{R}^3} x_i \cdot \frac{d}{dx_i} \left(\frac{d\psi}{dx_i} \cdot \bar{\psi} - \frac{d\bar{\psi}}{dx_i} \cdot \psi \right) dx \\ &= -\frac{ih}{2m} \int_{\mathbb{R}^3} \left(\frac{d\psi}{dx_i} \cdot \bar{\psi} - \frac{d\bar{\psi}}{dx_i} \cdot \psi \right) dx \\ &= -\frac{ih}{m} \int_{\mathbb{R}^3} \frac{d\psi}{dx_i} \cdot \bar{\psi} \, dx \end{aligned}$$

(4) We conclude that the average speed is given by the following formula:

$$< v >= -\frac{ih}{m} \int_{\mathbb{R}^3} \nabla\psi \cdot \bar{\psi} \, dx$$

By multiplying by the mass, we obtain the formula for $< p >$ in the statement. □

As an interesting speculation now, based on the above two formulae, and inspired from Heisenberg's idea of matrix mechanics, we have:

SPECULATION 2.15. *The average position and momentum formulae, written as*

$$< x >= \int_{\mathbb{R}^3} \bar{\psi} \cdot x \cdot \psi \, dx$$

$$< p >= \int_{\mathbb{R}^3} \bar{\psi} \cdot (-ih\nabla) \cdot \psi \, dx$$

suggest that x represents position, and $-ih\nabla$ represents momentum.

To be more precise, here we don't quite know what the quantities x and $-ih\nabla$ really are, mathematically speaking, so let us call them for the moment "operators", and we will see later for axioms. We will discuss this, axioms, in chapter 3 below.

The point now is that, with this convention, the above speculation tells us that for computing the average value of the position and momentum x, p, we must "sandwich" the corresponding operator between $\bar{\psi}, \psi$, and then integrate.

Which is something quite remarkable, and we are now very tempted to formulate something extremely general, and of course still a bit vague, as follows:

SPECULATION 2.16. *The average value of an observable O should appear as*

$$< O >= \int_{\mathbb{R}^3} \bar{\psi} \cdot \widehat{O} \cdot \psi \, dx$$

"sandwich between $\bar{\psi}, \psi$ and integrate", where $\widehat{O}$ is the operator associated to O.

As an illustration, let us see if this sandwiching method works for the kinetic energy of the particle. The kinetic energy is given by the following formula:

$$T = \frac{m||v||^2}{2} = \frac{< p, p >}{2m}$$

Thus, the operator associated to the energy should be given by:

$$\widehat{T} = \frac{< -ih\nabla, -ih\nabla >}{2m} = -\frac{h^2\Delta}{2m}$$

We obtain in this way something which looks quite reasonable, as follows:

$$< T >= -\frac{h^2}{2m} \int_{\mathbb{R}^3} \Delta\psi \cdot \bar{\psi} \, dx$$

We will see later, in chapter 3 below, more explanations on all this.

More generally now, we can incorporate into our method the potential energy too, and we are led in this way to the following interesting, conceptual conclusion:

CONCLUSION 2.17. *According to the above speculations, the operator associated to the total energy, or Hamiltonian, $H = T + V$ is given by*

$$\widehat{H} = -\frac{h^2\Delta}{2m} + V$$

and so the Schrödinger equation itself appears as

$$ih\dot{\psi} = \widehat{H}\psi$$

in terms of this operator, as claimed in Comment 2.11.

To be more precise, according to the above, $\widehat{H}$ appears indeed via the formula in the statement. But now, let us look back at the Schrödinger equation, namely:

$$ih\dot{\psi} = -\frac{h^2}{2m}\Delta\psi + V\psi$$

We recognize on the right the operator $\widehat{H}$ acting on ψ, and we are led to the conclusion in the statement. But probably enough for now on this topic, and more later.

Back to computations now, and to the Schrödinger equation as it is, simple and clear equation, let us investigate the case of time-independent potentials, as is the case of the Coulomb potential of the proton, that we are mostly interested in. We have here:

THEOREM 2.18. *In the case of time-independent potentials V, which include the Coulomb potential of the proton, the solutions of the Schrödinger equation*

$$ih\dot{\psi} = -\frac{h^2}{2m}\Delta\psi + V\psi$$

which are of the following special form, with the time and space variables separated,

$$\psi_t(x) = w_t\phi(x)$$

are given by the following formulae, with E being a certain constant,

$$w = e^{-iEt/h}w_0 \quad , \quad E\phi = -\frac{h^2}{2m}\Delta\phi + V\phi$$

with the equation for ϕ being called time-independent Schrödinger equation.

PROOF. This follows indeed by doing some math, as follows:

(1) Assuming that we have $\psi = w\phi$ as in the statement, we obtain:

$$\dot{\psi} = \dot{w}\phi \quad , \quad \Delta\psi = w\Delta\phi$$

Thus, the Schrödinger equation reformulates as follows:

$$ih\dot{w}\phi = -\frac{h^2}{2m}w\Delta\phi + Vw\phi$$

By dividing now everything by $w\phi$, our equation becomes:

$$ih\cdot\frac{\dot{w}}{w}=-\frac{h^2}{2m}\cdot\frac{\Delta\phi}{\phi}+V$$

(2) Now observe that the left-hand side depends only on time, and the right-hand side depends only on space. Thus, we must have, for a certain constant E:

$$ih\cdot\frac{\dot{w}}{w}=-\frac{h^2}{2m}\cdot\frac{\Delta\phi}{\phi}+V=E$$

(3) Let us first examine the first equation, involving time, namely:

$$ih\cdot\frac{\dot{w}}{w}=E$$

This equation can be written more conveniently as follows:

$$\frac{d}{dt}\log w=-\frac{iE}{h}$$

Thus we have $w=e^{-iEt/h}w_0$, as claimed in the statement.

(4) Regarding now the second equation, involving space, this is:

$$-\frac{h^2}{2m}\cdot\frac{\Delta\phi}{\phi}+V=E$$

But by multiplying by ϕ, this gives the second equation in the statement. □

As a first remark, the above makes the link with the speculations from Comment 2.10, and we can now formulate, as a complement to Conclusion 2.17:

Conclusion 2.19. *The Schrödinger equation naturally appears from the de Broglie claim on the wave properties of the electron, as claimed in Comment 2.10.*

This is something very nice, and together with Conclusion 2.17, it brings a more conceptual point of view on the Schrödinger equation. We will be back to all this in a moment, when talking axiomatization, based on these facts.

As a second comment, the above results, when coupled with some extra computations, show that the electron is not a particle in the classical sense, the reason being that a classical particle wave function cannot satisfy the time-independent Schrödinger equation. Thus, to put it squarely, in connection with the considerations from the previous section, the harm to Newton is there, in the Schrödinger approach, but hidden well under the carpet. More on this later, when talking about axiomatization.

2b.

2c.

2d.

2e. Exercises

Exercises:

Exercise 2.20.

Exercise 2.21.

Exercise 2.22.

Exercise 2.23.

Exercise 2.24.

Exercise 2.25.

Exercise 2.26.

Exercise 2.27.

Bonus exercise.

CHAPTER 3

Quantum mechanics

3a. Quantum mechanics

We discuss here the axiomatization of quantum mechanics, following Heisenberg and Schrödinger, and then Dirac and others. Hang on, tough material to come.

We already talked in chapter 1 about the main idea of Heisenberg, namely using infinite matrices in order to axiomatize quantum mechanics, based on the following key fact, coming from the discoveries of Balmer, and then Lyman, Paschen and others:

FACT 3.1 (Rydberg, Ritz). *The spectral lines of the hydrogen atom are given by the Rydberg formula, as follows, depending on integer parameters $n_1 < n_2$:*

$$\frac{1}{\lambda_{n_1n_2}} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

These spectral lines combine according to the Ritz-Rydberg principle, as follows:

$$\frac{1}{\lambda_{n_1n_2}} + \frac{1}{\lambda_{n_2n_3}} = \frac{1}{\lambda_{n_1n_3}}$$

Similar formulae hold for other atoms, with suitable fine-tunings of the constant R.

We refer to chapter 1 for the full story with all this, which is theory based on some key observations of Lyman, Balmer, Paschen, around 1890-1900. The point now is that the above combination principle reminds the multiplication formula $e_{n_1n_2}e_{n_2n_3} = e_{n_1n_3}$ for the elementary matrices $e_{ij} : e_j \to e_i$, which leads to the following principle:

PRINCIPLE 3.2 (Heisenberg). *Observables in quantum mechanics should be some sort of infinite matrices, generalizing the Lyman, Balmer, Paschen lines of the hydrogen atom, and multiplying between them as the matrices do, as to produce further observables.*

All this is quite deep, and needs a number of comments, as follows:

(1) First of all, our matrices must be indeed infinite, because so are the series observed by Lyman, Balmer, Paschen, corresponding to $n_1 = 1, 2, 3$ in the Rydberg formula, and making it clear that the range of the second parameter $n_2 > n_1$ is up to ∞.

(2) Although this was not known to Ritz-Rydberg and Heisenberg, let us mention too that some later results of Brackett, Pfund, Humphreys and others, at $n_1 = 4, 5, 6, \ldots$, confirmed the fact that the range of the first parameter n_1 is up to ∞ too.

(3) As a more tricky comment now, going beyond what Principle 3.2 says, our infinite matrices must be in fact complex. This was something known to Heisenberg, and later Schrödinger came with proof that quantum mechanics naturally lives over $\mathbb{C}$.

(4) But all this leads us into some tricky mathematics, because the infinite matrices $A \in M_\infty(\mathbb{C})$ do not act on the vectors $v \in \mathbb{C}^\infty$ just like that. For instance the all-one matrix $A_{ij} = 1$ does not act on the all-one vector $v_i = 1$, for obvious reasons.

Summarizing, in order to get to some mathematical theory going, out of Principle 3.2, we must assume that our matrices $A \in M_\infty(\mathbb{C})$ must be "bounded" in some sense. Or perhaps the vectors $v \in \mathbb{C}^\infty$ must be bounded. Or perhaps, both.

In order to fix all this, let us start with the questions regarding the space $\mathbb{C}^\infty$. In view of the above, we would like to replace it with its subspace $H = l^2(\mathbb{N})$ consisting of vectors having finite norm, as for our various computations to converge.

This being said, taking a look at what Schrödinger was saying too, a bit later, why not including right away in our theory spaces like $H = L^2(\mathbb{R}^3)$ too, which are perhaps a bit more relevant than Heisenberg's $l^2(\mathbb{N})$. We are led in this way into:

DEFINITION 3.3. *A Hilbert space is a complex vector space H with a scalar product $< x, y >$, which will be linear at left and antilinear at right,*

$$< \lambda x, y >= \lambda < x, y > \quad , \quad < x, \lambda y >= \bar{\lambda} < x, y >$$

and which is complete with respect to corresponding norm

$$||x|| = \sqrt{< x, x >}$$

in the sense that any sequence $\{x_n\}$ which is a Cauchy sequence, having the property $||x_n - x_m|| \to 0$ with $n, m \to \infty$, has a limit, $x_n \to x$.

Observe that we are using here the mathematicians' notation $< ., . >$ and convention for the scalar products, with these being linear at left. There are several reasons for preferring this notation, and more on this later, trust me in the meantime.

Getting now to work, there is some mathematics encapsulated in the above definition, certainly needing some discussion. First, we have the following result:

THEOREM 3.4. *Given an index set I, which can be finite or not, the space of square-summable vectors having indices in I, namely*

$$l^2(I) = \left\{ (x_i)_{i\in I} \Big| \sum_i |x_i|^2 < \infty \right\}$$

is a Hilbert space, with scalar product as follows:

$$< x, y >= \sum_i x_i \bar{y}_i$$

When I is finite, $I = \{1, \ldots, N\}$, we obtain in this way the usual space $H = \mathbb{C}^N$.

PROOF. This can be done in several steps, as follows:

(1) Given a vector $x \in \mathbb{C}^I$, let us define its norm by the following formula:

$$||x|| = \sqrt{\sum_i |x_i|^2}$$

We know that $l^2(I) \subset \mathbb{C}^I$ is the space of vectors satisfying $||x|| < \infty$. We want to prove that $l^2(I)$ is a vector space, that $< x, y >$ is a scalar product on it, that $l^2(I)$ is complete with respect to $||.||$, and finally that for $|I| < \infty$ we have $l^2(I) = \mathbb{C}^{|I|}$.

(2) The last assertion, $l^2(I) = \mathbb{C}^{|I|}$ for $|I| < \infty$, is clear, because in this case the sums are finite, so the condition $||x|| < \infty$ is automatic. So, we know at least one thing.

(3) Regarding the rest, our claim here, which will more or less prove everything, is that for any two vectors $x, y \in l^2(I)$ we have the Cauchy-Schwarz inequality:

$$| < x, y > | \leq ||x|| \cdot ||y||$$

(4) In order to prove this inequality, consider the following quantity, depending on a real variable $t \in \mathbb{R}$, and on a variable on the unit circle, $w \in \mathbb{T}$:

$$f(t) = ||twx + y||^2$$

By developing f, we see that this is a degree 2 polynomial in t:

$$\begin{aligned} f(t) &= < twx + y, twx + y > \\ &= t^2 < x, x > +tw < x, y > +t\bar{w} < y, x > + < y, y > \\ &= t^2||x||^2 + 2tRe(w < x, y >) + ||y||^2 \end{aligned}$$

Since f is obviously positive, its discriminant must be negative:

$$4Re(w < x, y >)^2 - 4||x||^2 \cdot ||y||^2 \leq 0$$

But this is equivalent to the following condition:

$$|Re(w < x, y >)| \leq ||x|| \cdot ||y||$$

Now the point is that we can arrange for the number $w \in \mathbb{T}$ to be such that the quantity $w < x, y >$ is real. Thus, we obtain, as desired:

$$| < x, y > | \leq ||x|| \cdot ||y||$$

(5) As a side remark here, observe that the equality case happens precisely when the discriminant of f vanishes, so when f has a root, and so when x, y are proportional.

(6) Now with Cauchy-Schwarz proved, everything is straightforward. We first obtain, by raising to the square and expanding, that for any $x, y \in l^2(I)$ we have:

$$||x + y|| \leq ||x|| + ||y||$$

Thus $l^2(I)$ is indeed a vector space, the other vector space conditions being trivial.

(7) Also, $< x, y >$ is surely a scalar product on this vector space, because all the conditions for a scalar product, which are as follows, are satisfied:

$* < x, y >$ is linear in x, and antilinear in y.

$* \overline{< x, y >} =< y, x >$, for any x, y.

$* < x, x >> 0$, for any $x \neq 0$.

(8) Finally, the fact that our space $l^2(I)$ is indeed complete with respect to its norm $||.||$ follows in the obvious way, the limit of a Cauchy sequence $\{x_n\}$ being the vector $y = (y_i)$ given by $y_i = \lim_{n\to\infty} x_{ni}$, with all the verifications here being trivial. □

Going now a bit abstract, we have, more generally, the following result, which shows that our formalism covers as well the Schrödinger spaces of type $L^2(\mathbb{R}^3)$:

THEOREM 3.5. *Given an arbitrary space X with a positive measure μ on it, the space of square-summable complex functions on it, namely*

$$L^2(X) = \left\{ f : X \to \mathbb{C} \middle| \int_X |f(x)|^2 \, d\mu(x) < \infty \right\}$$

is a Hilbert space, with scalar product as follows:

$$< f, g >= \int_X f(x)\overline{g(x)} \, d\mu(x)$$

When $X = I$ is discrete, meaning that the measure μ on it is the counting measure, $\mu(\{x\}) = 1$ for any $x \in X$, we obtain in this way the previous spaces $l^2(I)$.

PROOF. This is something routine, remake of Theorem 3.5, as follows:

(1) The proof of the first, and main assertion is something perfectly similar to the proof of Theorem 3.5, by replacing everywhere the sums by integrals.

(2) With the remark that we forgot to say in the statement that the L^2 functions are by definition taken up to equality almost everywhere, $f = g$ when $||f - g|| = 0$.

(2) As for the last assertion, when μ is the counting measure all our integrals here become usual sums, and so we recover in this way Theorem 3.5. □

As a third and last theorem about Hilbert spaces, that we will need, we have:

THEOREM 3.6. *Any Hilbert space H has an orthonormal basis $\{e_i\}_{i\in I}$, which is by definition a set of vectors whose span is dense in H, and which satisfy*

$$< e_i, e_j >= \delta_{ij}$$

with δ being a Kronecker symbol. The cardinality $|I|$ of the index set, which can be finite, countable, or worse, depends only on H, and is called dimension of H. We have

$$H \simeq l^2(I)$$

in the obvious way, mapping $\sum \lambda_i e_i \to (\lambda_i)$. The Hilbert spaces with $\dim H = |I|$ being countable, including $l^2(\mathbb{N})$ and $L^2(\mathbb{R})$, are all isomorphic, and are called separable.

PROOF. We have many assertions here, the idea being as follows:

(1) In finite dimensions an orthonormal basis $\{e_i\}_{i\in I}$ can be constructed by starting with any vector space basis $\{x_i\}_{i\in I}$, and using the Gram-Schmidt procedure. As for the other assertions, these are all clear, from basic linear algebra.

(2) In general, the same method works, namely Gram-Schmidt, with one subtlety coming from the fact that the basis $\{e_i\}_{i\in I}$ will not span in general the whole H, but just a dense subspace of it, as it is in fact obvious by looking for instance at the standard basis of $l^2(\mathbb{N})$. And there is a second subtlety as well, coming from the fact that the recurrence procedure needed for Gram-Schmidt must be replaced by some sort of "transfinite recurrence", using scary tools from logic, and more specifically the Zorn lemma.

(3) Finally, everything at the end is clear from definitions, except perhaps for the fact that $L^2(\mathbb{R})$ is separable. But here we can argue that, since functions can be approximated by polynomials, we have a countable algebraic basis, namely $\{x^n\}_{n\in\mathbb{N}}$, called the Weierstrass basis, that we can orthogonalize afterwards by using Gram-Schmidt. □

Observe that, in contrast to Theorem 3.5 and Theorem 3.6, there are several non-trivial things going on with Theorem 3.7. First we have the full proof of the basis existence, based on the Zorn lemma, which normally takes 1-2 pages, but which can easily take 5-6 pages, if you really want that Zorn lemma proved too, that we have of course avoided. But then, we have also some subtleties at the end, with the space $L^2(\mathbb{R})$ being in theory separable, but in practice not really, because the orthogonalization of the Weierstrass basis $\{x^n\}_{n\in\mathbb{N}}$ is something quite complicated. More on this later.

Moving ahead, now that we know what our vector spaces are, we can talk about infinite matrices with respect to them. Again, this will take some time.

Let us start with something elementary, as follows:

THEOREM 3.7. *For a linear operator $T : H \to H$, the following are equivalent:*

(1) *T is continuous.*
(2) *T is continuous at 0.*
(3) *T maps the unit ball of H into something bounded.*
(4) *T is bounded, in the sense that $||T|| = \sup_{||x||=1} ||Tx||$ is finite.*

PROOF. Here the equivalences (1) $\iff$ (2) $\iff$ (3) $\iff$ (4) all follow from definitions, by using the linearity of T, and performing various rescalings, and with the number $||T||$ needed in (4) being the bound coming from (3). □

With the above result in hand, we can now formulate:

THEOREM 3.8. *Given a Hilbert space H, consider the linear operators $T : H \to H$, and for each such operator define its norm by the following formula:*

$$||T|| = \sup_{||x||=1} ||Tx||$$

The operators which are bounded, $||T|| < \infty$, form then a complex algebra $B(H)$, which is complete with respect to $||.||$. When H comes with a basis $\{e_i\}_{i\in I}$, we have

$$B(H) \subset \mathcal{L}(H) \subset M_I(\mathbb{C})$$

where $\mathcal{L}(H)$ is the algebra of all linear operators $T : H \to H$, and $\mathcal{L}(H) \subset M_I(\mathbb{C})$ is the correspondence $T \to M$ obtained via the usual linear algebra formulae, namely:

$$T(x) = Mx \quad , \quad M_{ij} = < Te_j, e_i >$$

In infinite dimensions, none of the above two inclusions is an equality.

PROOF. This is something straightforward, the idea being as follows:

(1) The fact that we have indeed an algebra, satisfying the product condition in the statement, follows from the following estimates, which are all elementary:

$$||S + T|| \leq ||S|| + ||T||$$

$$||\lambda T|| = |\lambda| \cdot ||T||$$

$$||ST|| \leq ||S|| \cdot ||T||$$

(2) Regarding now the completness assertion, if $\{T_n\} \subset B(H)$ is Cauchy then $\{T_n x\}$ is Cauchy for any $x \in H$, so we can define the limit $T = \lim_{n\to\infty} T_n$ by setting:

$$Tx = \lim_{n\to\infty} T_n x$$

Let us first check that the application $x \to Tx$ is linear. We have:

$$\begin{aligned} T(x+y) &= \lim_{n\to\infty} T_n(x+y) \\ &= \lim_{n\to\infty} T_n(x) + T_n(y) \\ &= \lim_{n\to\infty} T_n(x) + \lim_{n\to\infty} T_n(y) \\ &= T(x) + T(y) \end{aligned}$$

Similarly, we have $T(\lambda x) = \lambda T(x)$, and we conclude that $T \in \mathcal{L}(H)$.

(3) With this done, it remains to prove now that we have $T \in B(H)$, and that $T_n \to T$ in norm. For this purpose, observe that we have:

$$\begin{aligned} ||T_n - T_m|| \leq \varepsilon \ , \ \forall n, m \geq N &\implies ||T_n x - T_m x|| \leq \varepsilon \ , \ \forall ||x|| = 1 \ , \ \forall n, m \geq N \\ &\implies ||T_n x - Tx|| \leq \varepsilon \ , \ \forall ||x|| = 1 \ , \ \forall n \geq N \\ &\implies ||T_N x - Tx|| \leq \varepsilon \ , \ \forall ||x|| = 1 \\ &\implies ||T_N - T|| \leq \varepsilon \end{aligned}$$

But this gives both $T \in B(H)$, and $T_N \to T$ in norm, and we are done.

(4) Regarding the embeddings, the correspondence $T \to M$ in the statement is indeed linear, and its kernel is $\{0\}$, so we have indeed an embedding as follows, as claimed:

$$\mathcal{L}(H) \subset M_I(\mathbb{C})$$

In finite dimensions we have an isomorphism, because any $M \in M_N(\mathbb{C})$ determines an operator $T : \mathbb{C}^N \to \mathbb{C}^N$, given by $< Te_j, e_i >= M_{ij}$. However, in infinite dimensions, we have matrices not producing operators, as for instance the all-one matrix.

(5) As for the examples of linear operators which are not bounded, these are more complicated, coming from logic, and we will not need them in what follows. □

Finally, as a thord and last result regarding the operators, we will need:

THEOREM 3.9. *Each operator $T \in B(H)$ has an adjoint $T^* \in B(H)$, given by:*

$$< Tx, y >=< x, T^* y >$$

The operation $T \to T^*$ is antilinear, antimultiplicative, involutive, and satisfies:

$$||T|| = ||T^*|| \quad , \quad ||TT^*|| = ||T||^2$$

When H comes with a basis $\{e_i\}_{i\in I}$, the operation $T \to T^*$ corresponds to

$$(M^*)_{ij} = \overline{M}_{ji}$$

at the level of the associated matrices $M \in M_I(\mathbb{C})$.

PROOF. This is standard too, and can be proved in 3 steps, as follows:

(1) The existence of the adjoint operator T^*, given by the formula in the statement, comes from the fact that the function $\varphi(x) =< Tx, y >$ being a linear map $H \to \mathbb{C}$, we must have a formula as follows, for a certain vector $T^*y \in H$:

$$\varphi(x) =< x, T^*y >$$

Moreover, since this vector is unique, T^* is unique too, and we have as well:

$$(S+T)^* = S^* + T^*$$

$$(\lambda T)^* = \bar{\lambda} T^*$$

$$(ST)^* = T^* S^*$$

$$(T^*)^* = T$$

Observe also that we have indeed $T^* \in B(H)$, because:

$$\begin{aligned} ||T|| &= \sup_{||x||=1} \sup_{||y||=1} < Tx, y > \\ &= \sup_{||y||=1} \sup_{||x||=1} < x, T^*y > \\ &= ||T^*|| \end{aligned}$$

(2) Regarding now $||TT^*|| = ||T||^2$, which is a key formula, observe that we have:

$$||TT^*|| \leq ||T|| \cdot ||T^*|| = ||T||^2$$

On the other hand, we have as well the following estimate:

$$\begin{aligned} ||T||^2 &= \sup_{||x||=1} | < Tx, Tx > | \\ &= \sup_{||x||=1} | < x, T^*Tx > | \\ &\leq ||T^*T|| \end{aligned}$$

By replacing $T \to T^*$ we obtain from this $||T||^2 \leq ||TT^*||$, as desired.

(3) Finally, when H comes with a basis, the formula $< Tx, y >=< x, T^*y >$ applied with $x = e_i$, $y = e_j$ gives the formula $(M^*)_{ij} = \overline{M}_{ji}$ in the statement. □

So, this was for the basics of operator theory, extending the basics of linear algebra. For more on all this, including full proofs for certain things in the above, you can check any book labeled functional analysis, or operator theory, or operator algebras, with a good reference here being the functional analysis book by Lax [66].

We are now ready for axiomatizing quantum mechanics. Following Heisenberg and Schrödinger, and then especially Dirac, who did the axiomatization work, we have:

AXIOMS 3.10. *In quantum mechanics the states of the system are vectors of a Hilbert space H, and the observables of the system are linear operators*

$$T : H \to H$$

which can be densely defined, and are taken self-adjoint, $T = T^*$. The average value of such an observable T, evaluated on a state $\xi \in H$, is given by:

$$< T >=< T\xi, \xi >$$

In the context of the Schrödinger mechanics of the hydrogen atom, the Hilbert space is the space $H = L^2(\mathbb{R}^3)$ where the wave function ψ lives, and we have

$$< T >= \int_{\mathbb{R}^3} T(\psi) \cdot \bar{\psi}\, dx$$

which is our previous "sandwiching" formula, with the operators

$$x \quad , \quad -\frac{ih}{m}\nabla \quad , \quad -ih\nabla \quad , \quad -\frac{h^2\Delta}{2m} \quad , \quad -\frac{h^2\Delta}{2m} + V$$

representing the position, speed, momentum, kinetic energy, and total energy.

In other words, we are doing here two things. First, we are declaring by axiom that our previous "sandwiching" formula holds true, and with this having all sorts of interesting consequences, already discussed before. And second, we are raising the possibility for other quantum mechanical systems, more complicated, to be described as well by the mathematics of the operators on a certain Hilbert space H, as above.

As a first result of our new theory, we have:

THEOREM 3.11 (Heisenberg). *We have the following uncertainty principle,*

$$\sigma_S \cdot \sigma_T \geq \left| \frac{< [S, T] >}{2} \right|$$

regarding the variances of any two observables S, T. In particular, we have

$$\sigma_x \cdot \sigma_p \geq \frac{h}{2}$$

implying that you cannot measure position and momentum at the same time.

PROOF. This follows indeed by doing some mathematics with operators and their commutators, and for details here, we refer for instance to Griffiths [43]. □

The above uncertainty principle, which is as old as quantum mechanics, is something quite surprising, that you can love or not. There are two schools of thought here, of Bohr and Einstein, and for more on all this, you can check the book by Kumar [59].

3b.

3c.

3d.

3e. Exercises

Exercises:

Exercise 3.12.

Exercise 3.13.

Exercise 3.14.

Exercise 3.15.

Exercise 3.16.

Exercise 3.17.

Exercise 3.18.

Exercise 3.19.

Bonus exercise.

CHAPTER 4

Hydrogen atom

4a. Hydrogen atom

Moving ahead towards hydrogen, we are interested in the case where V is the usual quadratic Coulomb potential of the proton, given by the following formula:

$$V = -\frac{e^2}{4\pi\varepsilon_0} \cdot \frac{1}{r}$$

This potential is obviously time-independent, and our main tool here will be the theory of the time-independent Schrödinger equation, as developed in the previous chapters. Let us recall from there that we have the following key result:

THEOREM 4.1. *In the case of time-independent potentials V, including the Coulomb potential of the proton, the solutions of the Schrödinger equation*

$$ih\dot{\psi} = -\frac{h^2}{2m}\Delta\psi + V\psi$$

which are of the following special form, with the time and space variables separated,

$$\psi_t(x) = w_t\phi(x)$$

are given by the following formulae, with E being a certain constant,

$$w = e^{-iEt/h}w_0 \quad , \quad E\phi = -\frac{h^2}{2m}\Delta\phi + V\phi$$

with the equation for ϕ being called time-independent Schrödinger equation.

PROOF. This is something that we know from chapter 2, that we have included here, for convenience. Let us recall too the idea. By dividing by ψ, the equation becomes:

$$ih \cdot \frac{\dot{w}}{w} = -\frac{h^2}{2m} \cdot \frac{\Delta\phi}{\phi} + V$$

Now since the left-hand side depends only on time, and the right-hand side depends only on space, both quantities must equal a constant E, and this gives the result. □

Moving ahead with theory, we can further build on Theorem 4.1, with a number of key observations on the time-independent Schrödinger equation, as follows:

THEOREM 4.2. *In the case of time-independent potentials V, the Schrödinger equation and its time-independent version have the following properties:*

(1) *For solutions of type $\psi = w_t\phi(x)$, the density $\varphi = |\psi|$ is time-independent, and more generally, all quantities of type $< T >$ are time-independent.*
(2) *The time-independent Schrödinger equation can be written as $\widehat{H}\phi = E\phi$, with $H = T + V$ being the total energy, of Hamiltonian.*
(3) *For solutions of type $\psi = w_t\phi(x)$ we have $< H^k >= E^k$ for any k. In particular we have $< H >= E$, and the variance is $< H^2 > - < H >^2 = 0$.*

PROOF. All the formulae are clear indeed from the fact that, when using the sandwiching formula for computing averages, the phases will cancel:

$$< T >= \int_{\mathbb{R}^3} \bar{\psi} \cdot T \cdot \psi \, dx = \int_{\mathbb{R}^3} \bar{\phi} \cdot T \cdot \phi \, dx$$

Thus, we are led to the various conclusions in the statement. □

All the above is quite interesting, physically speaking, and for a discussion here, we refer to Griffiths [43]. We will be back as well to this, a bit later.

We have as well the following result, mathematical this time:

THEOREM 4.3. *The solutions of the Schrödinger equation with time-independent potential V appear as linear combinations of separated solutions*

$$\psi = \sum_n c_n e^{-iE_n t/h} \phi_h$$

with the absolute values of the coefficients being given by

$$< H >= \sum_n |c_n|^2 E_n$$

$|c_n|$ being the probability for a measurement to return the energy value E_n.

PROOF. This is something standard, which follows from Fourier analysis, which allows the decomposition of ψ as in the statement, and that we will not really need, in what follows next. As before, for a physical discussion here, we refer to Griffiths [43]. □

Finally, a word about time-dependent potentials too, that we will ignore in this chapter. These are very important, due to the following:

REMARK 4.4. *For more complicated situations, like the helium atom, or heavier, the potential V in the Schrödinger equation is time-dependent, because the electron is subject here to the repulsions from the other electrons, which move in time.*

More on such potentials later, when taking helium and other atoms. In what follows we will be exclusively obsessed by hydrogen, where the math is simpler, and that we want to solve, above everything, anyway. By the way our obsession reminds that of the astrophysicists, who often call anything different from hydrogen "metals".

Moving ahead towards hydrogen, let us assume that V is the usual quadratic Coulomb potential of the proton. This potential is rotationally invariant, and it is convenient to use spherical coordinates, which are as follows, with $s \in [0, \pi]$ and $t \in [0, 2\pi]$:

$$\begin{cases} x &= r\cos s \\ y &= r\sin s\cos t \\ z &= r\sin s\sin t \end{cases}$$

We first must reformulate the Schrödinger equation in spherical coordinates. And for this purpose, we will need a well-known, scary computation, as follows:

THEOREM 4.5. *The Laplace operator in spherical coordinates is:*

$$\Delta = \frac{1}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d}{dr}\right) + \frac{1}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d}{ds}\right) + \frac{1}{r^2\sin^2 s}\cdot\frac{d^2}{dt^2}$$

PROOF. There are several proofs here, a short, elementary one being as follows:

(1) Let us first see how Δ behaves under a change of coordinates $\{x_i\} \to \{y_i\}$, in arbitrary N dimensions. Our starting point is the chain rule for derivatives:

$$\frac{d}{dx_i} = \sum_j \frac{d}{dy_j}\cdot\frac{dy_j}{dx_i}$$

By using this rule, then Leibnitz for products, then again this rule, we obtain:

$$\begin{aligned} \frac{d^2f}{dx_i^2} &= \sum_j \frac{d}{dx_i}\left(\frac{df}{dy_j}\cdot\frac{dy_j}{dx_i}\right) \\ &= \sum_j \frac{d}{dx_i}\left(\frac{df}{dy_j}\right)\cdot\frac{dy_j}{dx_i} + \frac{df}{dy_j}\cdot\frac{d}{dx_i}\left(\frac{dy_j}{dx_i}\right) \\ &= \sum_j \left(\sum_k \frac{d}{dy_k}\cdot\frac{dy_k}{dx_i}\right)\left(\frac{df}{dy_j}\right)\cdot\frac{dy_j}{dx_i} + \frac{df}{dy_j}\cdot\frac{d^2y_j}{dx_i^2} \\ &= \sum_{jk} \frac{d^2f}{dy_k dy_j}\cdot\frac{dy_k}{dx_i}\cdot\frac{dy_j}{dx_i} + \sum_j \frac{df}{dy_j}\cdot\frac{d^2y_j}{dx_i^2} \end{aligned}$$

(2) Now by summing over i, we obtain the following formula, with A being the derivative of $x \to y$, that is to say, the matrix of partial derivatives dy_i/dx_j:

$$\begin{aligned}\Delta f &= \sum_{ijk} \frac{d^2 f}{dy_k dy_j} \cdot \frac{dy_k}{dx_i} \cdot \frac{dy_j}{dx_i} + \sum_{ij} \frac{df}{dy_j} \cdot \frac{d^2 y_j}{dx_i^2} \\ &= \sum_{ijk} A_{ki} A_{ji} \frac{d^2 f}{dy_k dy_j} + \sum_{ij} \frac{d^2 y_j}{dx_i^2} \cdot \frac{df}{dy_j} \\ &= \sum_{jk} (AA^t)_{jk} \frac{d^2 f}{dy_k dy_j} + \sum_j \Delta(y_j) \frac{df}{dy_j}\end{aligned}$$

(3) So, this will be the formula that we will need. Observe that this formula can be further compacted as follows, with all the notations being self-explanatory:

$$\Delta f = Tr(AA^t H_y(f)) + < \Delta(y), \nabla_y(f) >$$

(4) Getting now to spherical coordinates, $(x, y, z) \to (r, s, t)$, the derivative of the inverse, obtained by differentiating x, y, z with respect to r, s, t, is given by:

$$A^{-1} = \begin{pmatrix} \cos s & -r \sin s & 0 \\ \sin s \cos t & r \cos s \cos t & -r \sin s \sin t \\ \sin s \sin t & r \cos s \sin t & r \sin s \cos t \end{pmatrix}$$

The product $(A^{-1})^t A^{-1}$ of the transpose of this matrix with itself is then:

$$\begin{pmatrix} \cos s & \sin s \cos t & \sin s \sin t \\ -r \sin s & r \cos s \cos t & r \cos s \sin t \\ 0 & -r \sin s \sin t & r \sin s \cos t \end{pmatrix} \begin{pmatrix} \cos s & -r \sin s & 0 \\ \sin s \cos t & r \cos s \cos t & -r \sin s \sin t \\ \sin s \sin t & r \cos s \sin t & r \sin s \cos t \end{pmatrix}$$

But everything simplifies here, and we have the following remarkable formula, which by the way is something very useful, worth to be memorized:

$$(A^{-1})^t A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 s \end{pmatrix}$$

Now by inverting, we obtain the following formula, in relation with the above:

$$AA^t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/r^2 & 0 \\ 0 & 0 & 1/(r^2 \sin^2 s) \end{pmatrix}$$

(5) Let us compute now the Laplacian of r, s, t. We first have the following formula, that we will use many times in what follows, and is worth to be memorized:

$$\begin{aligned}\frac{dr}{dx} &= \frac{d}{dx}\sqrt{x^2+y^2+z^2}\\ &= \frac{1}{2}\cdot\frac{2x}{\sqrt{x^2+y^2+z^2}}\\ &= \frac{x}{r}\end{aligned}$$

Of course the same computation works for y, z too, and we therefore have:

$$\frac{dr}{dx}=\frac{x}{r}\quad,\quad\frac{dr}{dy}=\frac{y}{r}\quad,\quad\frac{dr}{dz}=\frac{z}{r}$$

(6) By using the above formulae, twice, we can compute the Laplacian of r:

$$\begin{aligned}\Delta(r) &= \Delta\left(\sqrt{x^2+y^2+z^2}\right)\\ &= \frac{d}{dx}\left(\frac{x}{r}\right)+\frac{d}{dy}\left(\frac{y}{r}\right)+\frac{d}{dz}\left(\frac{z}{r}\right)\\ &= \frac{r^2-x^2}{r^3}+\frac{r^2-y^2}{r^3}+\frac{r^2-z^2}{r^3}\\ &= \frac{2}{r}\end{aligned}$$

(7) In what regards now s, the computation here goes as follows:

$$\begin{aligned}\Delta(s) &= \Delta\left(\arccos\left(\frac{x}{r}\right)\right)\\ &= \frac{d}{dx}\left(-\frac{\sqrt{r^2-x^2}}{r^2}\right)+\frac{d}{dy}\left(\frac{xy}{r^2\sqrt{r^2-x^2}}\right)+\frac{d}{dz}\left(\frac{xz}{r^2\sqrt{r^2-x^2}}\right)\\ &= \frac{2x\sqrt{r^2-x^2}}{r^4}+\frac{r^2(z^2-2y^2)+2x^2y^2}{r^4\sqrt{r^2-x^2}}+\frac{r^2(y^2-2z^2)+2x^2z^2}{r^4\sqrt{r^2-x^2}}\\ &= \frac{2x\sqrt{r^2-x^2}}{r^4}+\frac{x(2x^2-r^2)}{r^4\sqrt{r^2-x^2}}\\ &= \frac{x}{r^2\sqrt{r^2-x^2}}\\ &= \frac{\cos s}{r^2\sin s}\end{aligned}$$

(8) Finally, in what regards t, the computation here goes as follows:

$$\begin{aligned}\Delta(t) &= \Delta\left(\arctan\left(\frac{z}{y}\right)\right)\\ &= \frac{d}{dx}(0)+\frac{d}{dy}\left(-\frac{z}{y^2+z^2}\right)+\frac{d}{dz}\left(\frac{y}{y^2+z^2}\right)\\ &= 0-\frac{2yz}{(y^2+z^2)^2}+\frac{2yz}{(y^2+z^2)^2}\\ &= 0\end{aligned}$$

(9) We can now plug the data from (4) and (6,7,8) in the general formula that we found in (2) above, and we obtain in this way:

$$\begin{aligned}\Delta f &= \frac{d^2f}{dr^2}+\frac{1}{r^2}\cdot\frac{d^2f}{ds^2}+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}+\frac{2}{r}\cdot\frac{df}{dr}+\frac{\cos s}{r^2\sin s}\cdot\frac{df}{ds}\\ &= \frac{2}{r}\cdot\frac{df}{dr}+\frac{d^2f}{dr^2}+\frac{\cos s}{r^2\sin s}\cdot\frac{df}{ds}+\frac{1}{r^2}\cdot\frac{d^2f}{ds^2}+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}\\ &= \frac{1}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{df}{dr}\right)+\frac{1}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{df}{ds}\right)+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}\end{aligned}$$

Thus, we are led to the formula in the statement. □

We can now reformulate the Schrödinger equation in spherical coordinates, and then separate the variables, which leads to a radial and angular equation, as follows:

THEOREM 4.6. *The time-independent Schrödinger equation in spherical coordinates separates, for solutions of type $\phi=\rho(r)\alpha(s,t)$, into two equations, as follows,*

$$\frac{d}{dr}\left(r^2\cdot\frac{d\rho}{dr}\right)-\frac{2mr^2}{h^2}(V-E)\rho=K\rho$$

$$\sin s\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\alpha}{ds}\right)+\frac{d^2\alpha}{dt^2}=-K\sin^2 s\cdot\alpha$$

with K being a constant, called radial equation, and angular equation.

PROOF. By using the formula in Theorem 4.5, the time-independent Schrödinger equation reformulates in spherical coordinates as follows:

$$(V-E)\phi=\frac{h^2}{2m}\left[\frac{1}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d\phi}{dr}\right)+\frac{1}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\phi}{ds}\right)+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2\phi}{dt^2}\right]$$

Let us look now for separable solutions for this latter equation, consisting of a radial part and an angular part, as in the statement, namely:

$$\phi(r,s,t)=\rho(r)\alpha(s,t)$$

By plugging this function into our equation, we obtain:

$$(V-E)\rho\alpha=\frac{h^2}{2m}\left[\frac{\alpha}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d\rho}{dr}\right)+\frac{\rho}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\alpha}{ds}\right)+\frac{\rho}{r^2\sin^2 s}\cdot\frac{d^2\alpha}{dt^2}\right]$$

In order to solve this equation, we will do two manipulations. First, by multiplying everything by $2mr^2/(h^2\rho\alpha)$, this equation takes the following more convenient form:

$$\frac{2mr^2}{h^2}(V-E)=\frac{1}{\rho}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d\rho}{dr}\right)+\frac{1}{\alpha\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\alpha}{ds}\right)+\frac{1}{\alpha\sin^2 s}\cdot\frac{d^2\alpha}{dt^2}$$

Now observe that by moving the radial terms to the left, and the angular terms to the right, this latter equation can be written as follows:

$$\frac{2mr^2}{h^2}(V-E)-\frac{1}{\rho}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d\rho}{dr}\right)=\frac{1}{\alpha\sin^2 s}\left[\sin s\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\alpha}{ds}\right)+\frac{d^2\alpha}{dt^2}\right]$$

Since this latter equation is now separated between radial and angular variables, both sides must be equal to a certain constant $-K$, as follows:

$$\frac{2mr^2}{h^2}(V-E)-\frac{1}{\rho}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d\rho}{dr}\right)=-K$$

$$\frac{1}{\alpha\sin^2 s}\left[\sin s\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\alpha}{ds}\right)+\frac{d^2\alpha}{dt^2}\right]=-K$$

But this leads to the conclusion in the statement. □

Let us first study the angular equation, and this for reasons that will become clear later, the idea being that this equation forces the constant K to be of the form $K=l(l+1)$ with $l\in\mathbb{N}$, which can be used afterwards in the study of the radial equation.

The study will be quite long. We first have the following result:

PROPOSITION 4.7. *The angular equation that we found before, namely*

$$\sin s\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\alpha}{ds}\right)+\frac{d^2\alpha}{dt^2}=-K\sin^2 s\cdot\alpha$$

separates, for solutions of type $\alpha=\sigma(s)\theta(t)$, into two equations, as follows,

$$\frac{1}{\theta}\cdot\frac{d^2\theta}{dt^2}=-m^2$$

$$\frac{\sin s}{\sigma}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\sigma}{ds}\right)+K\sin^2 s=m^2$$

with m being a constant, called azimuthal equation, and polar equation.

PROOF. This is something elementary, the idea being as follows:

(1) Before anything, for such questions, we need to have a better understanding of the angles s, t, and the differences between them. So, recall that these angles come from:

$$\begin{cases} x &= r\cos s \\ y &= r\sin s\cos t \\ z &= r\sin s\sin t \end{cases}$$

To be more precise, here $r \in [0,\infty)$ is the radius, $s \in [0,\pi]$ is the polar angle, and $t \in [0,2\pi]$ is the azimuthal angle. Be said in passing, there are several conventions and notations here, and the above ones, that we use here, come from the general ones in N dimensions, because further coordinates can be easily added, in the obvious way.

(2) Getting back now to our question, by plugging $\alpha = \sigma(s)\theta(t)$ into the angular equation, we obtain:

$$\sin s\cdot\theta\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\sigma}{ds}\right)+\sigma\cdot\frac{d^2\theta}{dt^2}=-K\sin^2 s\cdot\sigma\theta$$

By dividing everything by $\sigma\theta$, this equation can be written as follows:

$$-\frac{1}{\theta}\cdot\frac{d^2\theta}{dt^2}=\frac{\sin s}{\sigma}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\sigma}{ds}\right)+K\sin^2 s$$

Since the variables are separated, we must have, for a certain constant m:

$$\frac{1}{\theta}\cdot\frac{d^2\theta}{dt^2}=-m^2$$

$$\frac{\sin s}{\sigma}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\sigma}{ds}\right)+K\sin^2 s=m^2$$

Thus, we are led to the conclusion in the statement. □

Regarding the azimuthal equation, things here are quickly settled, as follows:

PROPOSITION 4.8. *The solutions of the azimuthal equation, namely*

$$\frac{1}{\theta}\cdot\frac{d^2\theta}{dt^2}=-m^2$$

are the functions as follows, with $a, b \in \mathbb{C}$ being parameters,

$$\theta(t)=ae^{imt}+be^{-imt}$$

and with only the case $m \in \mathbb{Z}$ being acceptable, on physical grounds.

PROOF. The first assertion is clear, because we have a second order equation, and two obvious solutions for it, $e^{\pm imt}$, and then their linear combinations, and that's all. Regarding the last assertion, the point here is that by using $\theta(t)=\theta(t+2\pi)$, which is a natural physical assumption on the wave function, we are led to $m \in \mathbb{Z}$, as stated. □

We are now about to solve the angular equation, with only the polar equation remaining to be studied. However, in practice, this polar equation is 10 times more difficult that everything what we did so far, and so please be patient. We first have:

PROPOSITION 4.9. *The polar equation that we found before, namely*

$$\frac{\sin s}{\sigma}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\sigma}{ds}\right)+K\sin^2 s=m^2$$

with $m\in\mathbb{Z}$, translates via $\sigma(s)=f(\cos s)$ into the following equation,

$$(1-x^2)f''(x)-2xf'(x)=\left(\frac{m^2}{1-x^2}-K\right)f(x)$$

where $x=\cos s$, called Legendre equation.

PROOF. Let us first do a number of manipulations on our equation, before making the change of variables. By multiplying by σ, our equation becomes:

$$\sin s\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d\sigma}{ds}\right)=\left(m^2-K\sin^2 s\right)\sigma$$

By differentiating at left, this equation becomes:

$$\sin s\left(\cos s\cdot\sigma'+\sin s\cdot\sigma''\right)=\left(m^2-K\sin^2 s\right)\sigma$$

Finally, by dividing everything by $\sin^2 s$, our equation becomes:

$$\sigma''+\frac{\cos s}{\sin s}\cdot\sigma'=\left(\frac{m^2}{\sin^2 s}-K\right)\sigma$$

Now let us set $\sigma(s)=f(\cos s)$. With this change of variables, we have:

$$\sigma=f(\cos s)$$

$$\sigma'=-\sin s\cdot f'(\cos s)$$

$$\sigma''=-\cos s\cdot f'(\cos s)+\sin^2 s\cdot f''(\cos s)$$

By plugging this data, our radial equation becomes:

$$\sin^2 s\cdot f''(\cos s)-2\cos s\cdot f'(\cos s)=\left(\frac{m^2}{\sin^2 s}-K\right)f(\cos s)$$

Now with $x=\cos s$, which is our new variable, this equation reads:

$$(1-x^2)f''(x)-2xf'(x)=\left(\frac{m^2}{1-x^2}-K\right)f(x)$$

But this is the Legendre equation, as stated. □

Here comes now the difficult point. We have the following non-trivial result:

THEOREM 4.10. *The solutions of the Legendre equation, namely*

$$(1-x^2)f''(x)-2xf'(x)=\left(\frac{m^2}{1-x^2}-K\right)f(x)$$

can be explicitely computed, via complicated math, and only the case

$$K=l(l+1)\quad:\quad l\in\mathbb{N}$$

is acceptable, on physical grounds.

PROOF. The first part is something quite complicated, involving the hypergeometric functions ${}_2F_1$, that you don't want to hear about, believe me. As for the second part, analysis and physical speculations, this is something not trivial either. □

So, what to do? We will not fight with such extreme questions, and instead we will go very slowly, constructing from scratch the solutions which are "acceptable", with full details. And in what regards their uniqueness, well, we will refer to Theorem 4.10, whose proof can be certainly found somewhere, if you are really interested in that.

In order to construct the solutions, let us start with an extremely basic and fundamental problem. We have seen in chapter 3 that all Hilbert spaces of type $L^2(X)$ with $X\subset\mathbb{R}$ are separable, the reason behind this being the fact that we can start with the Weierstrass basis $\{x^l\}$, and then orthogonalize by Gram-Schmidt. However, as also mentioned in chapter 3, the Gram-Schmidt orthogonalization, while certainly being something that works in theory, is something quite complicated, if you want to do it explicitly.

Time now to understand this. For the simplest compact space $X\subset\mathbb{R}$, or unit ball of $\mathbb{R}$ if you prefer, which is the interval $[-1,1]$, this problem can be solved as follows:

THEOREM 4.11. *The orthonormal basis of $L^2[-1,1]$ obtained by starting with the Weierstrass basis $\{x^l\}$, and doing Gram-Schmidt, is the family of polynomials $\{P_l\}$, with each P_l being of degree l, and with positive leading coefficient, subject to:*

$$\int_{-1}^{1}P_k(x)P_l(x)\,dx=\delta_{kl}$$

These polynomials, called Legendre polynomials, satisfy the equation

$$(1-x^2)P_l''(x)-2xP_l'(x)+l(l+1)P_l(x)=0$$

which is the Legendre equation at $m=0$, and with $K=l(l+1)$. Moreover,

$$P_l(x)=\frac{1}{2^l l!}\left(\frac{d}{dx}\right)^l(x^2-1)^l$$

which is called the Rodrigues formula for Legendre polynomials.

PROOF. As a first observation, we are not lost somewhere in abstract math, because of the occurrence of the Legendre equation. As for the proof, this goes as follows:

(1) The first assertion is clear, because the Gram-Schmidt procedure applied to the Weierstrass basis $\{x^l\}$ can only lead to a certain family of polynomials $\{P_l\}$, with each P_l being of degree l, and also unique, if we assume that it has positive leading coefficient, with this $\pm$ choice being needed, as usual, at each step of Gram-Schmidt.

(2) In order to have now an idea about these beasts, here are the first few of them, which can be obtained say via a straightforward application of Gram-Schmidt:

$$\begin{aligned}
P_0 &= 1\\
P_1 &= x\\
P_2 &= (3x^2-1)/2\\
P_3 &= (5x^3-3x)/2\\
P_4 &= (35x^4-30x^2+3)/8\\
P_5 &= (63x^5-70x^3+15x)/8
\end{aligned}$$

(3) Now thinking about what Gram-Schmidt does, this is certainly something by recurrence. And examining the recurrence leads to the Legendre equation, as stated.

(4) As for the Rodrigues formula, by uniqueness no need to try to understand where this formula comes from, and we have two choices here, either by verifying that $\{P_l\}$ is orthonormal, or by verifying the Legendre equation. And both methods work. □

Going ahead now, we can solve in fact the Legendre equation at any m, as follows:

PROPOSITION 4.12. *The general Legendre equation, with parameters $m \in \mathbb{N}$ and $K = l(l+1)$ with $l \in \mathbb{N}$, namely*

$$(1-x^2)f''(x) - 2xf'(x) = \left(\frac{m^2}{1-x^2} - l(l+1)\right) f(x)$$

is solved by the following functions, called Legendre functions,

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \left(\frac{d}{dx}\right)^m P_l(x)$$

where P_l are as before the Legendre polynomials. Also, we have

$$P_l^m(x) = (-1)^m \frac{(1-x^2)^{m/2}}{2^l l!} \left(\frac{d}{dx}\right)^{l+m} (x^2-1)^l$$

called Rodrigues formula for Legendre functions.

PROOF. The first assertion is something elementary, coming by differentiating m times the Legendre equation, which leads to the general Legendre equation. As for the second assertion, this follows from the Rodrigues formula for Legendre polynomials. □

And this is the end of our study. Eventually. By putting together all the above results, the last 6 of them to be more precise, we are led to the following conclusion:

THEOREM 4.13. *The separated solutions $\alpha = \sigma(s)\theta(t)$ of the angular equation,*

$$\sin s \cdot \frac{d}{ds}\left(\sin s \cdot \frac{d\alpha}{ds}\right) + \frac{d^2\alpha}{dt^2} = -K \sin^2 s \cdot \alpha$$

are given by the following formulae, where $l \in \mathbb{N}$ is such that $K = l(l+1)$,

$$\sigma(s) = P_l^m(\cos s) \quad , \quad \theta(t) = e^{imt}$$

and where $m \in \mathbb{Z}$ is a constant, and with P_l^m being the Legendre function,

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \left(\frac{d}{dx}\right)^m P_l(x)$$

where P_l are the Legendre polynomials, given by the following formula:

$$P_l(x) = \frac{1}{2^l l!} \left(\frac{d}{dx}\right)^l (x^2-1)^l$$

These solutions $\alpha = \sigma(s)\theta(t)$ are called spherical harmonics.

PROOF. This follows indeed from all the above, and with the comment that everything is taken up to linear combinations. We will normalize the wave function later. □

As an inevitable comment here, our study of the angular equation was not complete, with all sorts of easy things missing, but also with something non-trivial not done, namely the uniqueness. This was discussed in Theorem 4.10, redirecting you towards ${}_2F_1$ and hypergeometric functions. Up to you here, depending on what you want to do in life. The point is that hypergeometric functions can depend on $N = 1, 2, 3, 4, 5, 6, 7, \ldots$ variables, and the bigger your N that you master, the better your math and physics will be.

As examples here, most scientists get away with $N = 1, 2$. A good mathematician must do $N = 3, 4$. As for good theoretical physics, $N = 5, 6, 7$ is usually required. For the story, I once had a difficult problem, and gave it to a $N = 7$ physicist, who solved it right away. And it took me a few good months to understand his solution.

In order now to finish our study, and eventually get to conclusions about hydrogen, it remains to solve the radial equation, for the Coulomb potential V of the proton.

Let us begin with some generalities, valid for any time-independent potential V. As a first manipulation on the radial equation, we have:

PROPOSITION 4.14. *The radial equation, written with $K = l(l+1)$,*

$$(r^2\rho')' - \frac{2mr^2}{h^2}(V-E)\rho = l(l+1)\rho$$

takes with $\rho = u/r$ the following form, called modified radial equation,

$$Eu = -\frac{h^2}{2m}\cdot u'' + \left(V + \frac{h^2 l(l+1)}{2mr^2}\right)u$$

which is a time-independent 1D Schrödinger equation.

PROOF. With $\rho = u/r$ as in the statement, we have:

$$\rho = \frac{u}{r} \quad , \quad \rho' = \frac{u'r - u}{r^2} \quad , \quad (r^2\rho')' = u''r$$

By plugging this data into the radial equation, this becomes:

$$u''r - \frac{2mr}{h^2}(V-E)u = \frac{l(l+1)}{r}\cdot u$$

By multiplying everything by $h^2/(2mr)$, this latter equation becomes:

$$\frac{h^2}{2m}\cdot u'' - (V-E)u = \frac{h^2 l(l+1)}{2mr^2}\cdot u$$

But this gives the formula in the statement. As for the interpretation, as time-independent 1D Schrödinger equation, this is clear as well, and with the comment here that the term added to the potential V is some sort of centrifugal term. □

Let us now, eventually, get to hydrogen. Here V is the usual quadratic Coulomb potential of the proton, given by the following formula, with e being as usual the charge of the electron, and ε_0 being the electric permittivity of free space:

$$V = -\frac{e^2}{4\pi\varepsilon_0}\cdot\frac{1}{r}$$

However, before getting into math, we must first discuss units. Remember from electrodynamics the story of the Coulomb constant K, which eventually gets replaced by $\varepsilon_0 = 1/(4\pi K)$, due to the Gauss law, and the Maxwell equations? Well, the Maxwell equations being now obsolete, not to say wrong, in quantum mechanics, time to welcome back the Coulomb constant K. Our new conventions will be as follows:

CONVENTIONS 4.15. *We welcome back the Coulomb constant K, given by:*

$$K = 8.987\ 551\ 7923(14) \times 10^9$$

Also, we welcome as new quantity for energy the electron volt eV,

$$1eV = e = 1.602\ 176\ 634 \times 10^{-19}$$

with this being regarded, as per our SI philosophy, as a constant, not a unit.

As usual, lots of fun here with units. In what regards the Coulomb constant K and minus the charge of the electron e, these are given by the formulae in the statement, with the formula of e being exact, as per latest SI regulations. As for the electron volt eV, this is by definition the amount of kinetic energy gained by an electron accelerating from rest through an electric potential difference of 1 volt in vacuum. Which in practice means that 1 eV is simply the number e, but regarded as a constant for energy.

Getting back now to the Coulomb potential of the proton, we have here:

FACT 4.16. *The Coulomb potential of the hydrogen atom proton, acting on the electron by attraction, is given according to the Coulomb law by*

$$V = -\frac{Kep}{r}$$

where p is the charge of the proton, and K is the Coulomb constant. In practice however we have $p \simeq e$ up to order 10^{-7}, and so our formula can be written as

$$V \simeq -\frac{Ke^2}{r}$$

and we will use this latter formula, and with $=$ sign, for simplifying.

Getting back now to math, it remains to solve the modified radial equation, for the above potential V. And we have here the following result, which does not exactly solve this radial equation, but provides us instead with something far better, namely the proof of the original claim by Bohr, which was at the origin of everything:

THEOREM 4.17 (Schrödinger). *In the case of the hydrogen atom, where V is the Coulomb potential of the proton, the modified radial equation, which reads*

$$Eu = -\frac{h^2}{2m} \cdot u'' + \left(-\frac{Ke^2}{r} + \frac{h^2 l(l+1)}{2mr^2} \right) u$$

leads to the Bohr formula for allowed energies,

$$E_n = -\frac{m}{2} \left(\frac{Ke^2}{h} \right)^2 \cdot \frac{1}{n^2}$$

with $n \in \mathbb{N}$, the binding energy being

$$E_1 \simeq -2.177 \times 10^{-18}$$

with means $E_1 \simeq -13.591$ eV.

PROOF. This is again something non-trivial, and we will be following Griffiths [43], with some details missing. The idea is as follows:

(1) By dividing our modified radial equation by E, this becomes:

$$-\frac{h^2}{2mE}\cdot u''=\left(1+\frac{Ke^2}{Er}-\frac{h^2l(l+1)}{2mEr^2}\right)u$$

In terms of $\alpha=\sqrt{-2mE}/h$, this equation takes the following form:

$$\frac{u''}{\alpha^2}=\left(1+\frac{Ke^2}{Er}+\frac{l(l+1)}{(\alpha r)^2}\right)u$$

In terms of the new variable $p=\alpha r$, this latter equation reads:

$$u''=\left(1+\frac{\alpha Ke^2}{Ep}+\frac{l(l+1)}{p^2}\right)u$$

Now let us introduce a new constant S for our problem, as follows:

$$S=-\frac{\alpha Ke^2}{E}$$

In terms of this new constant, our equation reads:

$$u''=\left(1-\frac{S}{p}+\frac{l(l+1)}{p^2}\right)u$$

(2) The idea will be that of looking for a solution written as a power series, but before that, we must "peel off" the asymptotic behavior. Which is something that can be done, of course, heuristically. With $p\to\infty$ we are led to $u''=u$, and ignoring the solution $u=e^p$ which blows up, our approximate asymptotic solution is:

$$u\sim e^{-p}$$

Similarly, with $p\to 0$ we are led to $u''=l(l+1)u/p^2$, and ignoring the solution $u=p^{-l}$ which blows up, our approximate asymptotic solution is:

$$u\sim p^{l+1}$$

(3) The above heuristic considerations suggest writing our function u as follows:

$$u=p^{l+1}e^{-p}v$$

So, let us do this. In terms of v, we have the following formula:

$$u'=p^le^{-p}\left[(l+1-p)v+pv'\right]$$

Differentiating a second time gives the following formula:

$$u''=p^le^{-p}\left[\left(\frac{l(l+1)}{p}-2l-2+p\right)v+2(l+1-p)v'+pv''\right]$$

Thus the radial equation, as modified in (1) above, reads:

$$pv''+2(l+1-p)v'+(S-2(l+1))v=0$$

(4) We will be looking for a solution v appearing as a power series:

$$v = \sum_{j=0}^{\infty} c_j p^j$$

But our equation leads to the following recurrence formula for the coefficients:

$$c_{j+1} = \frac{2(j+l+1) - S}{(j+1)(j+2l+2)} \cdot c_j$$

(5) We are in principle done, but we still must check that, with this choice for the coefficients c_j, our solution v, or rather our solution u, does not blow up. And the whole point is here. Indeed, at $j >> 0$ our recurrence formula reads, approximately:

$$c_{j+1} \simeq \frac{2c_j}{j}$$

But, surprisingly, this leads to $v \simeq c_0 e^{2p}$, and so to $u \simeq c_0 p^{l+1} e^p$, which blows up.

(6) As a conclusion, the only possibility for u not to blow up is that where the series defining v terminates at some point. Thus, we must have for a certain index j:

$$2(j+l+1) = S$$

In other words, we must have, for a certain integer $n > l$:

$$S = 2n$$

(7) We are almost there. Recall from (1) above that S was defined as follows:

$$S = -\frac{\alpha K e^2}{E} \quad : \quad \alpha = \frac{\sqrt{-2mE}}{h}$$

Thus, we have the following formula for the square of S:

$$S^2 = \frac{\alpha^2 K^2 e^4}{E^2} = -\frac{2mE}{h^2} \cdot \frac{K^2 e^4}{E^2} = -\frac{2mK^2 e^4}{h^2 E}$$

Now by using the formula $S = 2n$ from (6), the energy E must be of the form:

$$E = -\frac{2mK^2 e^4}{h^2 S^2} = -\frac{mK^2 e^4}{2h^2 n^2}$$

Calling this energy E_n, depending on $n \in \mathbb{N}$, we have, as claimed:

$$E_n = -\frac{m}{2}\left(\frac{Ke^2}{h}\right)^2 \cdot \frac{1}{n^2}$$

(8) Thus, we proved the Bohr formula. Regarding now the numerics, the data is as follows, with all formulae being of course approximative:

$$K = 8.988 \times 10^9 \quad , \quad e = 1.602 \times 10^{-19}$$

$$h = 1.055 \times 10^{-34} \quad , \quad m = 9.109 \times 10^{-31}$$

We obtain succesively that we have the following formulae:

$$\frac{Ke^2}{h} = \frac{8.988 \times 1.602^2}{1.055} \times \frac{10^9 \times 10^{-38}}{10^{-34}} = 2.186 \times 10^6$$

$$\left(\frac{Ke^2}{h}\right)^2 = 2.186^2 \times 10^{12} = 4.779 \times 10^{12}$$

$$\frac{m}{2}\left(\frac{Ke^2}{h}\right)^2 = \frac{9.109 \times 4.779}{2} \times 10^{12-31} = 2.177 \times 10^{-18}$$

Thus E_1 is as in the statement. In electron volts now, the figure is:

$$\frac{E_1}{e} = \frac{2.177 \times 10^{-18}}{1.602 \times 10^{-19}} = 13.591$$

Thus, we are led to the conclusion in the statement. □

As a first remark, all this agrees with the Rydberg formula, due to:

THEOREM 4.18. *The Rydberg constant for hydrogen is given by*

$$R = -\frac{E_1}{h_0 c}$$

where E_1 is the Bohr binding energy, and the Rydberg formula itself, namely

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

simply reads, via the energy formula in Theorem 4.17,

$$\frac{1}{\lambda_{n_1 n_2}} = \frac{E_{n_2} - E_{n_1}}{h_0 c}$$

which is in agreement with the Planck formula $E = h_0 c/\lambda$.

PROOF. Here the first assertion is something numeric, coming from the fact that the formula in the statement gives, when evaluated, the Rydberg constant:

$$R = \frac{-E_1}{h_0 c} = \frac{2.177 \times 10^{-18}}{6.626 \times 10^{-34} \times 2.998 \times 10^8} = 1.096 \times 10^7$$

As a consequence, and passed now what the experiments exactly say, we can define the Rydberg constant of hydrogen abstractly, by the following formula:

$$R = \frac{m}{2h_0 c}\left(\frac{Ke^2}{h}\right)^2$$

Regarding now the second assertion, by dividing $R = -E_1/(h_0 c)$ by any number of type n^2 we obtain, according to the energy convention in Theorem 4.17:

$$\frac{R}{n^2} = -\frac{E_n}{h_0 c}$$

But these are exactly the numbers which are subject to substraction in the Rydberg formula, and so we are led to the conclusion in the statement. □

With these spectacular applications explained, let us go back now to our study of the Schrödinger equation, done throughout this chapter. Our conclusions are:

THEOREM 4.19. *The wave functions of the hydrogen atom are the following functions, labelled by three quantum numbers, n, l, m,*

$$\phi_{nlm}(r, s, t) = \rho_{nl}(r)\alpha_l^m(s, t)$$

where $\rho_{nl}(r) = p^{l+1}e^{-p}v(p)/r$ with $p = \alpha r$ as before, with the coefficients of v subject to

$$c_{j+1} = \frac{2(j + l + 1 - n)}{(j + 1)(j + 2l + 2)} \cdot c_j$$

and $\alpha_l^m(s, t)$ being the spherical harmonics found before.

PROOF. This follows indeed by putting together all the results obtained so far, and with the remark that everything is up to the normalization of the wave function. □

In what regards the main wave function, that of the ground state, we have:

THEOREM 4.20. *With the hydrogen atom in its ground state, the wave function is*

$$\phi_{100}(r, s, t) = \frac{1}{\sqrt{\pi a^3}}\, e^{-r/a}$$

where $a = 1/\alpha$ is the inverse of the parameter appearing in our computations above,

$$\alpha = \frac{\sqrt{-2mE}}{h}$$

called Bohr radius of the hydrogen atom. This Bohr radius is the mean distance between the electron and the proton, in the ground state, and is given by the formula

$$a = \frac{h^2}{mKe^2}$$

which numerically means $a \simeq 5.291 \times 10^{-11}$.

PROOF. There are several things going on here, as follows:

(1) According to the various formulae in the proof of Theorem 4.7, taken at $n = 1$, the parameter α appearing in the computations there is given by:

$$\alpha = \frac{\sqrt{-2mE}}{h} = \frac{1}{h} \cdot m \cdot \frac{Ke^2}{h} = \frac{mKe^2}{h^2}$$

Thus, the inverse $\alpha = 1/a$ is indeed given by the formula in the statement.

(2) Regarding the wave function, according to Theorem 4.19 this consists of:

$$\rho_{10}(r) = \frac{2e^{-r/a}}{\sqrt{a^3}} \quad , \quad \alpha_0^0(s,t) = \frac{1}{2\sqrt{\pi}}$$

By making the product, we obtain the formula of ϕ_{100} in the statement.

(3) But this formula of ϕ_{100} shows in particular that the Bohr radius a is indeed the mean distance between the electron and the proton, in the ground state.

(4) Finally, in what regards the numerics, these are as follows:

$$a = \frac{1.055^2 \times 10^{-68}}{9.109 \times 10^{-31} \times 8.988 \times 10^9 \times 1.602^2 \times 10^{-38}} = 5.297 \times 10^{-11}$$

Thus, we are led to the conclusions in the statement. □

Getting back now to the general setting of Theorem 4.19, the point is that the polynomials $v(p)$ appearing there are well-known objects in mathematics, as follows:

PROPOSITION 4.21. *The polynomials $v(p)$ are given by the formula*

$$v(p) = L_{n-l-1}^{2l+1}(p)$$

where the polynomials on the right, called associated Laguerre polynomials, are given by

$$L_q^p(x) = (-1)^p \left(\frac{d}{dx}\right)^p L_{p+q}(x)$$

with L_{p+q} being the Laguerre polynomials, given by the following formula:

$$L_q(x) = \frac{e^x}{q!} \left(\frac{d}{dx}\right)^q (e^{-x} x^q)$$

PROOF. The story here is very similar to that of the Legendre polynomials. Consider the Hilbert space $H = L^2[0,\infty)$, with the following scalar product on it:

$$< f, g >= \int_0^\infty f(x) g(x) e^{-x}\, dx$$

(1) The orthogonal basis obtained by applying Gram-Schmidt to the Weierstrass basis $\{x^q\}$ is then the basis formed by the Laguerre polynomials $\{L_q\}$.

(2) We have the explicit formula for L_q in the statement, which is analogous to the Rodrigues formula for the Legendre polynomials.

(3) The first assertion follows from the fact that the coefficients of the associated Laguerre polynomials satisfy the equation for the coefficients of $v(p)$.

(4) Alternatively, the first assertion follows as well by using an equation for the Laguerre polynomials, which is very similar to the Legendre equation. □

With the above result in hand, we can now improve Theorem 4.19, as follows:

THEOREM 4.22. *The wave functions of the hydrogen atom are given by*

$$\phi_{nlm}(r,s,t)=\sqrt{\left(\frac{2}{na}\right)^3\frac{(n-l-1)!}{2n(n+l)!}}e^{-r/na}\left(\frac{2r}{na}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2r}{na}\right)\alpha_l^m(s,t)$$

with $\alpha_l^m(s,t)$ being the spherical harmonics found before.

PROOF. This follows indeed by putting together what we have, namely Theorem 4.19 and Proposition 4.21, and then doing some remaining work, concerning the normalization of the wave function, which leads to the normalization factor appearing above. □

And isn't this beautiful. If you want to impress your nerdy friends, or even a random customer in a pub, this is surely the formula that you want to show to them.

4b.

4c.

4d.

4e. Exercises

Exercises:

EXERCISE 4.23.

EXERCISE 4.24.

EXERCISE 4.25.

EXERCISE 4.26.

EXERCISE 4.27.

EXERCISE 4.28.

EXERCISE 4.29.

EXERCISE 4.30.

Bonus exercise.

Part II

Fine structure

When it all comes caving in
And you can't be brave again
When ever you need a friend
Need a friend, call me

CHAPTER 5

Electron spin

5a. Electron spin

We have seen that a theory of quantum mechanics can be developed, as to solve the hydrogen atom, at least approximately, along the lines suggested by Bohr. Our goal now will be to fine-tune these results, with a number of corrections that are needed, and also extend them to the case of heavier atoms, fully realizing Bohr's program.

We will of course need, learn and develop a lot of further quantum mechanics, of quite general type, in order to achieve this goal, following Pauli and Dirac, then Bose and Fermi, and many others. This further quantum mechanics can be of course used for doing all sorts of other things as well, and more on this later in this book.

In order to get started, the idea is that we have two basic corrections to our solution of the hydrogen atom. First we have a relativistic correction, which looks like something within our range, that can only be understood and computed. But then we have as well a spin-related correction, involving the notion of spin, which is totally new to us.

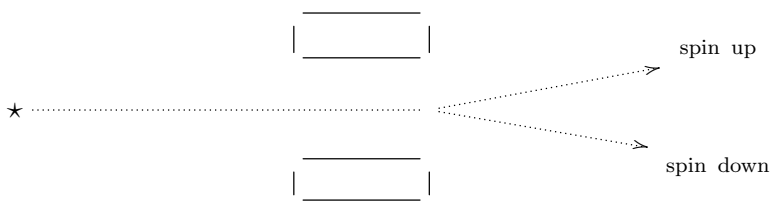
So, as a first question that we would like to solve, we have:

QUESTION 5.1. *What is the electron spin? That is, what experiments prove that the electron spins? And then, importantly, what is the mathematics of the spin?*

Talking mathematics first, the spin, if that beast exists indeed, is certainly not visible on the wave function ψ, because this wave function deals with position only. Thus, at least we know one thing, once the spin observed, we will most likely have to incorporate it into our theory by using the matrix mechanics formalism of Heisenberg.

Talking physics now, the main experiment leading to spin is as follows:

FACT 5.2 (Stern-Gerlach experiment). *When passing a beam of electrons through an inhomogeneous magnetic field, these electrons get deflected* $50-50$ *up or down,*



with the only possible explanation being that the electrons have a spin, which is $50-50$ *up or down. The same happens with a beam of neutral atoms, and a magnetic field strong enough, to be put at blame being the statistics of the spins of the constituents.*

So, this was the experiment, and what we call here "up" and "down" is of course the binary choice of the spin orientation, a bit as for usual, round objects in $\mathbb{R}^3$. That is, our Earth turns to the right, and in physics we would say that is has "spin up". Was the Earth turning to the left, we would say in physics that is has "spin down".

Of course, our presentation above is over-simplified. The original experiment was with neutral particles, namely silver atoms, and this in order to avoid the Lorentz force, which will curve the trajectory of any charged particle, to a much greater extent than the spin up/down deviation to be observed. Later experiments, with charged particles, used some extra apparatus, namely a suitable electric field, positioned after the electromagnet in the above diagram, designed to cancel the effects of the Lorentz force.

As an important observation, the Stern-Gerlach experiment does not observe the absolute, 3D spin up/down feature of the particles, but just a 1D component of it. However, it is possible to cascade experiments, by sending each of the output beams into separate Stern-Gerlach devices, and with these devices having various 3D orientations, and deduce some further conclusions from this. We refer here to Feynman [35].

So long for the Stern-Gerlach experiment. Getting back now to theory and speculations, as a first, innocent observation based on the above, we have:

OBSERVATION 5.3. *A single electron has an interesting life even when fixed, because it spins. Thus, no need for Heisenberg or Schrödinger for getting introduced to quantum mechanics, you can just try to understand the mathematics of a fixed electron.*

Moreover, as a cherry on the cake, as we will soon discover, the above-mentioned mathematics is that of the 2×2 complex matrices, which is at the same time something elementary, and fascinating. Which, getting us now into philosophy, leads us into the temptation of burying the physics, and talking right away about 2×2 matrices.

And shall we do this or not. Looking at the physics literature, there is a fair mess in the treatment of spin. At one end, you have spin-centered books, taking Observation 5.3 literally, and starting the book with a long, not to say never-ending, discussion about spin. Then you have quantum information related books, such as Bengtsson-Życzkowski [15], Nielsen-Chuang [67], Peres [70], which by a certain desire of brevity and efficiency, rapidly bury the physics of spin, and talk instead about 2×2 matrices. And then you have well-known and loved books such as Feynman [35], Griffiths [43], Weinberg [93], presenting all sorts of rather incomprehensible explanations regarding the spin, which vary with authors' taste, for eventually ending, of course, with 2×2 matrices.

And so again, what shall we do, talk about 2×2 matrices or not. Not clear. But, as usual in such difficult situations, we can always ask the cat. We haven't seen him in a while, but good news, cat is here, and willing to talk physics. And cat says:

CAT 5.4. *Be honest, and say what you have to say. And don't worry about your young readers, they will survive.*

This sounds wise as usual, thanks cat. So, we will follow this advice. But let me get first a huge mug of coffee, or rather huge mug of expresso, because fighting with the physics of the spin with bare hands is something which is reputed impossible.

To start with, and as a matter of reframing our discussion, and having something fresh to rely upon, let us demolish Observation 5.3 with:

FACT 5.5. *Observation 5.3 is something toxic. You can't really measure spin, and build a serious theory on that alone. What you need to do is to observe spin in context, via its tiny corrections to quantum mechanics. More specifically, spin is an order*

$$\alpha^2 \simeq \frac{1}{10,000}$$

correction to quantum mechanics, and more precisely to the Bohr energy formula, with the spin correction there appearing as a complement to the relativistic correction. And with this being the correct, healthy and constructive definition of spin.

In short, we are getting here back to the beginning, general quantum mechanics, with the main conclusion of the Stern-Gerlach experiment, namely "spin exists", recorded. Of course it is possible to say a bit more from Stern-Gerlach, namely recording the scattering angle, and doing some math there, but this basically does not advance us much. So better forget about Stern-Gerlach, and get back to general quantum mechanics.

The point now is that, with the above fact in hand, not only we are into truth, as we should be, but also we start getting an idea on how to reach to the mathematics of the spin. To be more precise, we should just think relativity, in the context of quantum mechanics, and with a bit of luck, all this thinking will lead us into spin.

In practice now, all this is doable, but a bit complicated, and was done by Klein, Gordon, Dirac a few years after Uhlenbeck, Goudsmit, Pauli came up with their theory of spin. So, let us briefly explain this idea, which is very beautiful, and we'll come later to Uhlenbeck, Goudsmit, Pauli. Consider the Schrödinger equation for a free electron:

$$ih\dot{\psi} = -\frac{h}{2m}\,\Delta\psi$$

Relativity theory dictates that the 3 space coordinates and the 1 time coordinate should be on the same footing, and so that we should replace $\dot{\psi}$ by something of type $\ddot{\psi}$. But this can be done by replacing the kinetic energy operator $T = \Delta/2m$ by its relativistic analogue, and also by invoking the invariance under Lorentz transformations, and we are led in this way to the following equation, called Klein-Gordon equation:

$$\left(\Delta - \frac{1}{c^2}\cdot\frac{d^2}{dt^2}\right)\psi = \frac{m^2c^2}{h^2}\,\psi$$

The point now, which is the key one, discovered by Dirac short after Klein and Gordon, is that it is possible to extract the square root of the Klein-Gordon operator:

$$\Delta - \frac{1}{c^2}\cdot\frac{d^2}{dt^2} = \left(\frac{Pd}{dx} + \frac{Qd}{dy} + \frac{Rd}{dz} + \frac{i}{c}\cdot\frac{Sd}{dt}\right)^2$$

Indeed, we need for this purpose matrices P, Q, R, S which anticommute, $AB = -BA$, and whose squares equal one, $A^2 = 1$. But such beasts can be found in $M_4(\mathbb{C})$, and then we can take the formal square root of the Klein-Gordon equation:

$$\left(\frac{Pd}{dx} + \frac{Qd}{dy} + \frac{Rd}{dz} + \frac{i}{c}\cdot\frac{Sd}{dt}\right)\psi = \frac{mc}{h}\,\psi$$

And the thing now, which is truly remarkable, is that this latter equation, called Dirac equation, does work indeed, in the sense that it is a true equation of physics, improving the Schrödinger equation. And a closer look at all this reveals afterwards that the fine structure of hydrogen, comprising the relativistic correction and the spin correction, can be understood in this way, leading to a clear mathematics of the spin.

All this is very beautiful, and leads us into:

THOUGHT 5.6. *Our criticism from Fact 5.5 was probably too harsh, relativity and spin alike being probably more than a mere*

$$\alpha^2 \simeq \frac{1}{10,000}$$

order correction to quantum mechanics. And this is because the Dirac equation, which is of first order, is something simpler than the Schrödinger equation.

In fact, we are now again into Observation 5.3, and this time armed with some solid math, and more specifically with a first-grade weapon, the Dirac equation. Which starts to be a bit tiring, yes I know, looks like we're changing our opinion about spin faster than Madonna is changing her shoes. But blame the cat, he came with his advice Cat 5.4.

Moving ahead now, and still following Cat 5.4, after some more thinking, the Dirac equation remains however something a bit speculative, or perhaps something too advanced, and it would be much better, at least to start with, to forget about relativity and abstractions, and have something more solid, regarding the spin.

And fortunately, there is a second way of viewing things, very elementary, inspired from our study of classical mechanics, or even from the movement of our good old Earth, which rotates and spins at the same time, which is as follows:

PHILOSOPHY 5.7. *In analogy with classical mechanics, spin should be something of same nature as angular momentum, coming on top of it.*

And good news, this will be our final, stable philosophy. Eventually.

To be more precise, following Uhlenbeck, Goudsmit, Pauli, we will first talk angular momentum, then we will axiomatize spin as being the quantity which naturally "complements" the angular momentum. Then we will talk about 2×2 matrices, and review the fine structure corrections to hydrogen as well. And finally, regarding the Klein-Gordon and Dirac equations, we will be back to them in chapter 7 below.

So, let us first talk about angular momentum, and we will get to spin later. We will need the following basic result, for doing computations:

PROPOSITION 5.8. *The components of the position operator $x = (x_1, x_2, x_3)$ and momentum operator $p = -ih\nabla$ satisfy the following relations,*

$$[x_i, x_j] = [p_i, p_j] = 0$$

$$[x_i, p_j] = ih\delta_{ij}$$

where $[a, b] = ab - ba$, called canonical commutation relations.

PROOF. All the above formulae are elementary, as follows:

(1) The components of the position operator $x = (x_1, x_2, x_3)$ obviously commute with each other, $x_i x_j = x_j x_i$, which makes their commutators vanish, $[x_i, x_j] = 0$.

(2) Regarding the momentum operator $p = -ih\nabla$, its components are as follows:

$$p_1 = -ih \cdot \frac{d}{dx_1} \quad , \quad p_2 = -ih \cdot \frac{d}{dx_2} \quad , \quad p_3 = -ih \cdot \frac{d}{dx_3}$$

Since partial derivatives commute with each other, we obtain $[p_i, p_j] = 0$.

(3) It remains to prove the last formula, and we have here:

$$\begin{aligned}[x_i,p_j]f &= (x_ip_j-p_jx_i)f\\ &= -ih\left(x_i\cdot\frac{df}{dx_j}-\frac{d}{dx_j}(x_if)\right)\\ &= -ih\left(x_i\cdot\frac{df}{dx_j}-\frac{dx_i}{dx_j}\cdot f-x_i\cdot\frac{df}{dx_j}\right)\\ &= ih\cdot\frac{dx_i}{dx_j}\cdot f\\ &= ih\delta_{ij}\cdot f\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

The above might look a bit complicated, and the simplest way to remember it is that "everyhing commutes", that is, $ab=ba$, except for the coordinates and momenta coordinates taken in the same direction, which are subject to the following rule:

$$x_ip_i=p_ix_i+ih$$

Getting now to angular momentum, it is convenient to change notation, with (x,y,z) instead of (x_1,x_2,x_3), due to the vector product involved, which will break the symmetry between coordinates. We have the following result, to start with:

THEOREM 5.9. *The components of the angular momentum operator*

$$L=x\times(-ih\nabla)$$

satisfy the following equations,

$$[L_x,L_y]=ihL_z$$
$$[L_y,L_z]=ihL_x$$
$$[L_z,L_x]=ihL_y$$

called commutation relations for the angular momentum.

PROOF. With the more familiar notation $p=-ih\nabla$ for momentum, or rather for the associated operator, the components of the angular momentum operator are:

$$L_x=yp_z-zp_y$$
$$L_y=zp_x-xp_z$$
$$L_z=xp_y-yp_x$$

Let us prove the first commutation relation. We have:

$$\begin{aligned}[L_x,L_y] &= [yp_z-zp_y,zp_x-xp_z]\\ &= [yp_z,zp_x]-[yp_z,xp_z]-[zp_y,zp_x]+[zp_y,xp_z]\end{aligned}$$

By heavily using the commutation relations from Proposition 5.8, we have:

$$[yp_z, zp_x] = yp_zzp_x - zp_xyp_z = y(zp_z - ih)p_x - zyp_xp_z = -ihyp_x$$

$$[yp_z, xp_z] = yp_zxp_z - xp_zyp_z = 0$$

$$[zp_y, zp_x] = zp_yzp_x - zp_xzp_y = 0$$

$$[zp_y, xp_z] = zp_yxp_z - xp_zzp_y = zxp_yp_z - x(zp_z - ih)p_y = ihxp_y$$

We conclude that the commutator that we were computing is given by the following formula, which is precisely the one in the statement:

$$\begin{aligned}[L_x, L_y] &= -ihyp_x + ihxp_y\\ &= ih(xp_y - yp_x)\\ &= ihL_z\end{aligned}$$

The proof of the other two commutation relations is similar, or can be simply obtained by invoking the cyclic invariance $x \to y \to z \to x$ of our problem, which cyclic invariance is not broken by the vector product $\times$ used, and so can indeed be invoked. □

As an interesting consequence of Theorem 5.9, we have:

PROPOSITION 5.10. *The following operator, called square of angular momentum*

$$L^2 = L_x^2 + L_y^2 + L_z^2$$

commutes with all 3 *operators* L_x, L_y, L_z.

PROOF. We have the following computation, to start with:

$$\begin{aligned}[L^2, L_x] &= (L_x^2 + L_y^2 + L_z^2)L_x - L_x(L_x^2 + L_y^2 + L_z^2)\\ &= L_y^2L_x + L_z^2L_x - L_xL_y^2 - L_xL_z^2\\ &= [L_y^2, L_x] + [L_z^2, L_x]\end{aligned}$$

The first commutator can be computed with a trick, as follows:

$$\begin{aligned}[L_y^2, L_x] &= L_yL_yL_x - L_xL_yL_y\\ &= L_yL_yL_x - L_yL_xL_y + L_yL_xL_y - L_xL_yL_y\\ &= L_y[L_y, L_x] + [L_y, L_x]L_y\\ &= L_y(-ihL_z) + (-ihL_z)L_y\\ &= -ih(L_yL_z + L_zL_y)\end{aligned}$$

The second commutator can be computed with the same trick, as follows:

$$\begin{aligned}[L_z^2, L_x] &= L_zL_zL_x - L_xL_zL_z\\ &= L_zL_zL_x - L_zL_xL_z + L_zL_xL_z - L_xL_zL_z\\ &= L_z[L_z, L_x] + [L_z, L_x]L_z\\ &= L_z(ihL_y) + (ihL_y)L_z\\ &= ih(L_zL_y + L_yL_z)\end{aligned}$$

Now by summing we obtain the following commutation relation, as desired:

$$[L^2, L_x] = 0$$

The proof of the other two commutation relations is similar, or we can simply invoke here the cyclic symmetry argument from the end of the proof of Theorem 5.9. □

Let us discuss now the diagonalization of the momentum operators L_x, L_y, L_z. Since these operators do not commute, we cannot hope for a joint diagonalization for them. Thus, we must choose one of them, and for reasons that will become clear later, when writing things in spherical coordinates, we will choose L_x.

In view of Proposition 5.10, this operator L_x does commute with L^2, and so we can hope for a joint diagonalization of L^2, L_x. And, so is what happens:

THEOREM 5.11. *The operators L^2, L_x diagonalize as*

$$L^2 f_l^m = h^2 l(l+1) f_l^m$$

$$L_x f_l^m = hm f_l^m$$

where $l \in \mathbb{N}/2$ and $m = -l, -l+1, \ldots, l-1, l$.

PROOF. This is something quite long, the idea being as follows:

(1) For reasons that will become clear later on, let us introduce two operators as follows, called raising and lowering operators:

$$L_+ = L_y + iL_z$$

$$L_- = L_y - iL_z$$

We will often deal with these operators at the same time, using the following notation:

$$L_\pm = L_y \pm iL_z$$

(2) We have the following computation:

$$\begin{aligned}[L_x, L_\pm] &= [L_x, L_y] \pm i[L_x, L_z] \\ &= ihL_z \pm i(-ihL_y) \\ &= h(iL_z \pm L_y) \\ &= \pm h(\pm iL_z + L_y) \\ &= \pm hL_\pm\end{aligned}$$

(3) Our claim now is that $L^2 f = \lambda f$, $L_x f = \mu f$ imply:

$$L^2(L_\pm f) = \lambda(L_\pm f)$$

$$L_x(L_\pm f) = (\mu \pm h)(L_\pm f)$$

Indeed, the first formula follows from:

$$\begin{aligned}L^2(L_\pm f) &= L_\pm(L^2 f) \\ &= L_\pm(\lambda f) \\ &= \lambda(L_\pm f)\end{aligned}$$

As for the second formula, this follows from:

$$\begin{aligned}L_x(L_\pm f) &= L_x L_\pm f \\ &= (L_x L_\pm - L_\pm L_x)f + L_\pm L_x f \\ &= \pm hL_\pm f + L_\pm(\mu f) \\ &= (\mu \pm h)(L_\pm f)\end{aligned}$$

(4) Now in view of the formulae found in (3), the raising and lowering operators act on the joint eigenfunctions of L^2, L_x, by leaving the L^2 eigenvalue unchanged, and by raising and lowering the eigenvalue of L_x. But both this raising process and lowering process for the eigenvalue of L_x cannot go on forever, because of the following estimate:

$$\begin{aligned}\lambda &= < L^2 > \\ &= < L_x^2 > + < L_y^2 > + < L_z^2 > \\ &= \mu^2 + < L_y^2 > + < L_z^2 > \\ &\geq \mu^2\end{aligned}$$

(5) In order to see exactly how the raising and lowering processes terminate, we will need some more computations. We have:

$$\begin{aligned} L_\pm L_\mp &= (L_y \pm iL_z)(L_y \mp iL_z) \\ &= L_y^2 + L_z^2 \mp i(L_yL_z - L_zL_y) \\ &= L_y^2 + L_z^2 \mp i(ihL_x) \\ &= L_y^2 + L_z^2 \pm hL_x \\ &= L^2 - L_x^2 \pm hL_x \end{aligned}$$

Thus, we have the following formula:

$$L^2 = L_\pm L_\mp + L_x^2 \mp hL_x$$

Now assuming $L_xf = hlf$, at termination of the raising process, we have:

$$\begin{aligned} L^2(f) &= (L_-L_+ + L_x^2 + hL_x)f \\ &= (0 + h^2l^2 + h^2l)f \\ &= h^2l(l+1)f \end{aligned}$$

Similarly, assuming $L_xf = hl'f$, at termination of the lowering process, we have:

$$\begin{aligned} L^2(f) &= (L_+ - L_- + L_x^2 - hL_x)f \\ &= (0 + h^2l'^2 - h^2l')f \\ &= h^2l'(l'-1)f \end{aligned}$$

Thus $l(l+1) = l'(l'-1)$, and since $l' = l+1$ is impossible, due to raising vs lowering, we must have $l' = -l$, and this leads to the conclusion in the statement.

(6) Finally, for being complete, the full and conceptual understanding of all the above imperatively requires a certain cat climbing a certain ladder, and for full details here, and for other things missing from the above proof, we refer to Griffiths [43]. □

Moving ahead now, still a lot of work left to be done. Our idea will be to write everything in spherical coordinates, and find the eigenfunctions. We have here:

THEOREM 5.12. *In spherical coordinates r, s, t we have*

$$L_x = -\frac{ih}{dt}$$

$$L_y = ih\left(\frac{\sin t}{ds} + \frac{\cos s\cos t}{\sin s}\cdot\frac{1}{dt}\right)$$

$$L_z = -ih\left(\frac{\cos t}{ds} - \frac{\cos s\sin t}{\sin s}\cdot\frac{1}{dt}\right)$$

and the spherical harmonics are joint eigenfunctions of L^2, L_x.

PROOF. We recall that, according to our usual, N-dimensional looking conventions, the spherical coordinates are as follows, with $r \in [0,\infty)$ being the radius, $s \in [0,\pi]$ being the polar angle, and $t \in [0,2\pi]$ being the azimuthal angle:

$$\begin{cases} x &= r\cos s \\ y &= r\sin s\cos t \\ z &= r\sin s\sin t \end{cases}$$

(1) We know that we have $L = -ihx \times \nabla$, so let us first compute ∇ in spherical coordinates. We have here, according to the chain rule for derivatives:

$$\begin{aligned} \nabla &= \begin{pmatrix} dr/dx & ds/dx & dt/dx \\ dr/dy & ds/dy & dt/dy \\ dr/dz & ds/dz & dt/dz \end{pmatrix} \begin{pmatrix} d/dr \\ d/ds \\ d/dt \end{pmatrix} \\ &= \begin{pmatrix} dx/dr & dy/dr & dz/dr \\ dx/ds & dy/ds & dz/ds \\ dx/dt & dy/dt & dz/dt \end{pmatrix}^{-1} \begin{pmatrix} d/dr \\ d/ds \\ d/dt \end{pmatrix} \end{aligned}$$

(2) On the other hand, we know from chapter 4 that we have:

$$\begin{pmatrix} dx/dr & dx/ds & dx/dt \\ dy/dr & dy/ds & dy/dt \\ dz/dr & dz/ds & dz/dt \end{pmatrix} = \begin{pmatrix} \cos s & -r\sin s & 0 \\ \sin s\cos t & r\cos s\cos t & -r\sin s\sin t \\ \sin s\sin t & r\cos s\sin t & r\sin s\cos t \end{pmatrix}$$

We also know from chapter 4 that this latter matrix, say A, satisfies:

$$A^tA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2 s \end{pmatrix}$$

Now if we call D the diagonal matrix on the right, we conclude that the matrix, say B, appearing in the above formula of ∇ is given by:

$$\begin{aligned} B &= (A^t)^{-1} \\ &= AD^{-1} \\ &= \begin{pmatrix} \cos s & -r\sin s & 0 \\ \sin s\cos t & r\cos s\cos t & -r\sin s\sin t \\ \sin s\sin t & r\cos s\sin t & r\sin s\cos t \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/r^2 & 0 \\ 0 & 0 & 1/(r^2\sin^2 s) \end{pmatrix} \\ &= \begin{pmatrix} \cos s & -\sin s/r & 0 \\ \sin s\cos t & \cos s\cos t/r & -\sin t/(r\sin s) \\ \sin s\sin t & \cos s\sin t/r & \cos t/(r\sin s) \end{pmatrix} \end{aligned}$$

(3) Thus, the angular momentum operator that we are looking for, $L = -ihx \times \nabla$, written more conveniently as $L = -ihx/r \times r\nabla$, is given by:

$$L = -ih \begin{pmatrix} \cos s \\ \sin s \cos t \\ \sin s \sin t \end{pmatrix} \times \begin{pmatrix} r\cos s & -\sin s & 0 \\ r\sin s\cos t & \cos s\cos t & -\sin t/\sin s \\ r\sin s\sin t & \cos s\sin t & \cos t/\sin s \end{pmatrix} \begin{pmatrix} d/dr \\ d/ds \\ d/dt \end{pmatrix}$$

And computing now the vector product gives the formula for L in the statement.

(4) Now with our explicit formula for L in hand, we next find that the raising and lowering operators are given by:

$$L_\pm = \pm he^{\pm it}\left(\frac{d}{ds} \pm i\,\frac{\cos s}{\sin s}\cdot\frac{1}{dt}\right)$$

Next, we find that these two operators satisfy the following formula:

$$L_+L_- = -h^2\left(\frac{d^2}{ds^2} + \frac{\cos s}{\sin s}\cdot\frac{d}{ds} + \frac{\cos^2 s}{\sin^2 s}\cdot\frac{d^2}{dt^2} + i\frac{d}{dt}\right)$$

And finally, by using this latter formula, we find that L^2 is given by:

$$L^2 = -h^2\left(\frac{1}{\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d}{ds}\right) + \frac{1}{\sin^2 s}\cdot\frac{d^2}{dt^2}\right)$$

(5) With all these formulae in hand, we can now finish. The eigenfunction equation for the above operator L^2, with eigenvalue $h^2l(l+1)$, is as follows:

$$-h^2\left(\frac{1}{\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d}{ds}\right) + \frac{1}{\sin^2 s}\cdot\frac{d^2}{dt^2}\right)f = h^2l(l+1)f$$

But this is precisely the angular equation from chapter 4. As for the eigenfunction equation for the operator L_x, with eigenvalue hm, this is as follows:

$$-\frac{ih}{dt}\,f = hmf$$

But this latter equation is equivalent to the azimuthal equation, also from chapter 4. Thus, we are dealing here with equations that we already know, and the solutions are the spherical harmonics that we found in chapter 4, as claimed. □

So long for angular momentum. And even more magic in a moment, when talking about spin. For more on all the above, we refer to Griffiths [43] or Weinberg [93].

In order to talk now about spin, we will regard, a bit as in the classical mechanics case, the spin and the angular momentum as being similar quantities. Thus, in analogy with the basic equations for angular momentum, we should have:

DEFINITION 5.13. *The components of the spin operator are subject to*

$$[S_x, S_y] = ihS_z$$

$$[S_y, S_z] = ihS_x$$

$$[S_z, S_x] = ihS_y$$

called commutation relations for the spin operator.

What we did here, with these axioms, is of course a bit heuristic. But this is quite reasonable, and for a more detailed version of the story, invoking rotational invariance as for getting to the above equations, for the angular momentum, spin, or any kind of "generalized angular momentum", in some reasonable sense, we refer for instance to Weinberg [93]. In what follows we will take Definition 5.13 as it is, and do some rotational invariance work later, in chapter 7 below, directly in the relativistic framework.

The point now is that, with the above relations in hand, which are identical to the commutation relations for the angular momentum, all the general results from the previous section, based on that commutation relations, extend to our present setting, simply by changing L into S everywhere. And in particular, we are led in this way to:

THEOREM 5.14. *We have the following diagonalization formulae*

$$S^2 f_s^m = h^2 s(s+1) f_s^m$$

$$S_x f_s^m = hm f_s^m$$

$$S_\pm f_s^m = h\sqrt{s(s+1) - m(m \pm 1)}\, f_s^{m\pm 1}$$

involving the operators $S^2 = S_x^2 + S_y^2 + S_z^2$, S_x and $S_\pm = S_y \pm iS_z$.

PROOF. Here the first two formulae are something that we already know, from the previous section, with L, j being replaced by S, s. As for the last formula, this is something that we did not need, in the L, j context, but that we will need now. We want to compute the constants $C_{s,\pm}^m$ making work the raising and lowering formula, namely:

$$S_\pm f_s^m = C_{s,\pm}^m f_s^{m\pm 1}$$

But this can be done by using $S^2 = S_\pm S_\mp + S_x^2 \mp hS_x$ and $S_\pm^* = S_\mp$, and we get:

$$C_{s,+}^m = h\sqrt{s(s+1) - m(m+1)}$$

$$C_{s,-}^m = h\sqrt{s(s+1) - m(m-1)}$$

Thus, we are led to the last formula in the statement, and we are done. □

In practice now, let us look for the simplest mathematical realization of spin. We know from the Stern-Gerlach experiment that the spin is something binary, that can be either up, or down. Thus, we are led, for fixed particles, to a quantum mechanics over $H = \mathbb{C}^2$, with spin up and down being represented by the following two vectors:

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

It remains now to see how the equations in Theorem 5.14 reformulate, in this $H = \mathbb{C}^2$ setting. But here, not many choices, and we are led to:

DEFINITION 5.15. *In the quantum mechanics of the spin, over $H = \mathbb{C}^2$, with*

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

being spin up and down, the spin is subject to the following equations, for $f = e_1, e_2$,

$$S^2 f = h^2 s(s+1) f$$

$$S_x f = h m_f f$$

$$S_\pm f = h\sqrt{s(s+1) - m_f(m_f \pm 1)}\, \check{f}$$

with parameters $s = 1/2$, $m_{e_1} = 1/2$, $m_{e_2} = -1/2$, and with $\{e_1, e_2\} = \{f, \check{f}\}$.

Here all the choices, and notably $s = 1/2$, are very natural in view of Theorem 5.14, because these are the choices providing a "minimal" realization of the equations in Theorem 5.14, in the smallest possible number of dimensions, namely $N = 2$. However, all this comes with a shade of mystery, or at least is not rock-solid enough as to be called theorem, and it is probably safer to use the term "definition", as we did above.

The point now is that the above questions can be solved, the result being:

THEOREM 5.16. *In the above $H = \mathbb{C}^2$ context, of the mechanics of a single, fixed electron, the components of the normalized spin $\sigma = 2S/h$ are as follows,*

$$\sigma_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad , \quad \sigma_y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_z = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

called Pauli matrices. In the general, dynamic context, where we already have a Hilbert space H for the wave function, spin can be introduced by using the space

$$H' = H \otimes \mathbb{C}^2$$

and using the above Pauli matrices for it, acting on the $\mathbb{C}^2$ part.

PROOF. The equations in Definition 5.15, written in full detail, are as follows:

$$S^2\begin{pmatrix}1\\0\end{pmatrix}=\frac{3h^2}{4}\begin{pmatrix}1\\0\end{pmatrix}\quad,\quad S^2\begin{pmatrix}0\\1\end{pmatrix}=\frac{3h^2}{4}\begin{pmatrix}0\\1\end{pmatrix}$$

$$S_x\begin{pmatrix}1\\0\end{pmatrix}=\frac{h}{2}\begin{pmatrix}1\\0\end{pmatrix}\quad,\quad S_x\begin{pmatrix}0\\1\end{pmatrix}=-\frac{h}{2}\begin{pmatrix}0\\1\end{pmatrix}$$

$$S_+\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}\quad,\quad S_+\begin{pmatrix}0\\1\end{pmatrix}=h\begin{pmatrix}1\\0\end{pmatrix}$$

$$S_-\begin{pmatrix}1\\0\end{pmatrix}=h\begin{pmatrix}0\\1\end{pmatrix}\quad,\quad S_-\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

Thus, we have the following formulae, for the various matrices involved:

$$S^2=\frac{3h^2}{4}\begin{pmatrix}1&0\\0&1\end{pmatrix}\quad,\quad S_x=\frac{h}{2}\begin{pmatrix}1&0\\0&-1\end{pmatrix}$$

$$S_+=h\begin{pmatrix}0&1\\0&0\end{pmatrix}\quad,\quad S_-=h\begin{pmatrix}0&0\\1&0\end{pmatrix}$$

In relation with what we want to prove, we have obtained the formula of S_x. Regarding now the formulae of S_y, S_z, these follow by solving the following system:

$$S_+=S_y+iS_z$$

$$S_-=S_y-iS_z$$

To be more precise, the computation for S_y goes as follows:

$$S_y=\frac{S_++S_-}{2}=\frac{h}{2}\begin{pmatrix}0&1\\1&0\end{pmatrix}$$

As for the computation for S_z, this goes as follows:

$$S_z=\frac{S_+-S_-}{2i}=\frac{h}{2i}\begin{pmatrix}0&1\\-1&0\end{pmatrix}=\frac{h}{2}\begin{pmatrix}0&-i\\i&0\end{pmatrix}$$

Thus, we are led to the conclusions in the statement. □

As a first consequence of the above, looking quite good, we have:

FACT 5.17. *Electrons have spin* $1/2$.

This is motivated of course by the formula $s=1/2$ in Definition 5.15, but this being said, at least from the perspective of what we know so far about electrons, this does not make much sense, logically speaking. Remember indeed that we're still living under the sword of Heisenberg's uncertainty principle, and so the electrons themselves, and therefore their spin too, remain rather mathematical objects, far away from concrete things like, say planets in the Solar system, turning around the Sun and spinning. And also, there is some unclarity with $1/2$ vs $\pm 1/2$, because the spin can be up or down.

This being said, some speculations are certainly possible. For instance the Pauli matrices all square up to one, $\sigma_i^2 = 1$, and a well-known interpretation of this is that "it takes 720° instead of the usual 360° to turn an electron back in place", leading to the conclusion that the spin of the electron is $360/720 = 1/2$.

In any case, $s = 1/2$ for the electron is good to know, and we will heavily use this formula in what follows, for all sorts of purposes. And we will talk about spin $\neq 1/2$ too, in Part III, with a general particle discussion, invloving bosons and fermions.

5b.

5c.

5d.

5e. Exercises

Exercises:

Exercise 5.18.

Exercise 5.19.

Exercise 5.20.

Exercise 5.21.

Exercise 5.22.

Exercise 5.23.

Exercise 5.24.

Exercise 5.25.

Bonus exercise.

CHAPTER 6

Fine structure

6a. Fine structure

What is next? All sorts of corrections to the Bohr formula that we found in chapter 4, due to various phenomena that we neglected in our computations, or rather in our modeling of the problem, which can be both of electric and relativistic nature.

But before getting into that, which is something quite technical, let us first enjoy what we found before in chapter 4, say by taking it as a final, exact result regarding the hydrogen atom. As a first conclusion, of quite philosophical nature, we have:

CONCLUSION 6.1. *The phenomenon of quantization appears, mathematically speaking, from certain equations which generically blow up, and force the various separation constants $C \in \mathbb{R}$ which appear to be integers, $C \in \mathbb{N}$.*

To be more precise, the phenomenon of quantization that we are talking about is of course the Bohr energy one, allowing discrete energies only, E_n with $n \in \mathbb{N}$, which is the mother of everything, in quantum mechanics. Looking back at the proof of this fact, separation constants $C \in \mathbb{R}$ which mysteriously became integers, $C \in \mathbb{N}$, was indeed the mathematical phenomenon behind this. Which appeared no less than 3 times:

(1) First when the azimuthal/polar separation parameter, denoted m^2, turned to be the square of an integer, $m \in \mathbb{Z}$.

(2) Then when the radial/angular separation constant K turned to be of a similar form, $K = l(l+1)$ with $l \in \mathbb{N}$.

(3) And finally in the context of the radial equation, where the parameter S there turned to be of the form $S = 2n$, with $n \in \mathbb{N}$.

This is very nice, we have now a clear mathematical idea about why things are quantized, in quantum mechanics. The 3 space coordinates and the 1 time coordinate, who usually live in peace, get into fights when it comes to differential equations.

As another comment now, in our study we dismissed several times all sorts of solutions, on various physical grounds, usually unacceptable blow up. But, at a more advanced level, some of these solutions make sense of course, due to the following fact:

FACT 6.2. *The hydrogen atom is not the general 2-body problem in quantum mechanics, but rather the case of confined, stable orbits. Some of the solutions which blow up correspond to scattering, in the context of an electron/proton meeting.*

Again, this is something a bit philosophical. In analogy with classical mechanics, what we did is to solve the planetary motion problem. But things like comets and asteroids still need to be investigated, for having a full theory. And which is something quite technical, called "scattering theory", that we will not get into here, in this book.

Back to work now, let us explain the series of corrections to the Schrödinger solution to the hydrogen atom. We will focus on energy only, so let us start by recalling:

THEOREM 6.3 (Schrödinger). *The energy of the ϕ_{nlm} state of the hydrogen atom is independent on the quantum numbers l, m, given by the Bohr formula*

$$E_n = -\frac{\alpha^2}{n^2} \cdot \frac{mc^2}{2}$$

where α is a dimensionless constant, called fine structure constant, given by

$$\alpha = \frac{Ke^2}{hc}$$

which in practice means $\alpha \simeq 1/137$.

PROOF. This is the Bohr energy formula that we know, proved by Schrödinger, and reformulated in terms of Sommerfeld's fine structure constant:

(1) We know from chapter 4 that we have the following formula, which can be written as in the statement, by using the fine structure constant α:

$$E_n = -\frac{m}{2}\left(\frac{Ke^2}{h}\right)^2 \cdot \frac{1}{n^2}$$

(2) Observe now that our modified Bohr formula can be further reformulated as follows, with T_c being the kinetic energy of the electron travelling at speed c:

$$E_n = -\frac{\alpha^2}{n^2} \cdot T_c$$

Thus α^2, and so α too, is dimensionless, as being a quotient of energies.

(3) Let us doublecheck however this latter fact, the check being instructive. With respect to the SI system that we use, the units for K, e, h, c are:

$$U_K = \frac{m^3 \cdot kg}{s^2 \cdot C^2} \quad , \quad U_e = C \quad , \quad U_h = \frac{m^2 \cdot kg}{s} \quad , \quad U_c = \frac{m}{s}$$

Thus the units for the fine structure constant α are, as claimed:

$$U_\alpha = U_C \cdot U_e^2 \cdot U_h^{-1} \cdot U_c^{-1} = \frac{m^3 \cdot kg}{s^2 \cdot C^2} \cdot C^2 \cdot \frac{s}{m^2 \cdot kg} \cdot \frac{s}{m} = 1$$

(4) In what regards now the numerics, these are as follows:

$$\alpha = \frac{Ke^2/h}{c} \simeq \frac{2.186 \times 10^6}{2.998 \times 10^8} = 7.291 \times 10^{-3} \simeq \frac{1}{137}$$

Here we have used a standard estimate for Ke^2/h, as in chapter 4. □

The fine structure constant α is a remarkable quantity, as obvious from the above, and more on it in a moment. Among its other magic features, it manages well 2π factors. Indeed, by using $K = 1/(4\pi\varepsilon_0)$ and $h = h_0/2\pi$, we can write this constant as:

$$\alpha = \frac{e^2}{2\varepsilon_0 h_0 c}$$

Finally, let us record the complete official data for α and its inverse α^{-1}:

$$\alpha = 0.007\ 297\ 352\ 5693(11)$$

$$\alpha^{-1} = 137.035\ 999\ 084(21)$$

As a final comment here, all this lengthy discussion about α might sound a bit like mania, or mysticism. But wait for it. Sometimes soon α will be part of your life.

Moving ahead now with corrections to Theorem 6.3, we will be very brief, and for further details, we refer as usual to our favorite books, Feynman [35], Griffiths [43] and Weinberg [93]. We first have the following result, which is something non-trivial:

THEOREM 6.4. *There is a relativistic correction to be made to the Bohr energy E_n of the state ϕ_{nlm}, depending on the quantum number l, given by*

$$\mathcal{E}_{nl} = \frac{\alpha^2 E_n}{n^2}\left(\frac{n}{l+1/2} - \frac{3}{4}\right)$$

coming by replacing the kinetic energy by the relativistic kinetic energy.

PROOF. According to Einstein, the relativistic kinetic energy is given by:

$$T = \frac{p^2}{2m} - \frac{p^4}{8m^3c^2} + \ldots$$

The Schrödinger equation, based on $T = p^2/2m$, must be therefore corrected with a $\mathcal{T} = -p^4/(8m^3c^2)$ term, and this leads to the above correction term $\mathcal{E}_{nl}$. □

Equally non-trivial is the following correction, independent from the above one:

THEOREM 6.5. *There is a spin-related correction to be made to the Bohr energy E_n of the state ϕ_{nlm}, depending on the number $j = l \pm 1/2$, given by*

$$\mathcal{E}_{nj} = -\frac{\alpha^2 E_n}{n^2} \cdot \frac{n(j-l)}{(l+1/2)(j+1/2)}$$

coming from the torque of the proton on the magnetic moment of the electron.

PROOF. As we explained in chapter 5, the electron has a spin $\pm 1/2$, which is naturally associated to the quantum number l, leading to the parameter $j = l \pm 1/2$. But, knowing now that the electron has a spin, the proton which moves around it certainly acts on its magnetic moment, and this leads to the above correction term $\mathcal{E}_{nj}$. □

So, these are the first two corrections to be made, and again, we refer to Feynman [35], Griffiths [43], Weinberg [93] for details. Obviously we don't quite know what we're doing here, but let us add now the above corrections to E_n, and see what we get. We obtain in this way one of the most famous formulae in quantum mechanics, namely:

THEOREM 6.6. *The energy levels of the hydrogen atom, taking into account the fine structure coming from the relativistic and spin-related correction, are given by*

$$E_{nj} = E_n \left[1 + \frac{\alpha^2}{n^2} \left(\frac{n}{j + 1/2} - \frac{3}{4} \right) \right]$$

with $j = l \pm 1/2$ being as above, and with α being the fine structure constant.

PROOF. We have the following computation, based on the above formulae:

$$\begin{aligned} \mathcal{E}_{nl} + \mathcal{E}_{nj} &= \frac{\alpha^2 E_n}{n^2} \left(\frac{n}{l + 1/2} - \frac{3}{4} - \frac{n(j - l)}{(l + 1/2)(j + 1/2)} \right) \\ &= \frac{\alpha^2 E_n}{n^2} \left(\frac{n}{l + 1/2} \left(1 - \frac{j - l}{j + 1/2} \right) - \frac{3}{4} \right) \\ &= \frac{\alpha^2 E_n}{n^2} \left(\frac{n}{j + 1/2} - \frac{3}{4} \right) \end{aligned}$$

Thus the corrected formula of the energy is as follows:

$$\begin{aligned} E_{nj} &= E_n + \mathcal{E}_{nl} + \mathcal{E}_{nj} \\ &= E_n + \frac{\alpha^2 E_n}{n^2} \left(\frac{n}{j + 1/2} - \frac{3}{4} \right) \end{aligned}$$

We are therefore led to the conclusion in the statement. □

Summarizing, quantum mechanics is more complicated than what originally appears from Schrödinger's solution of the hydrogen atom. Which was something quite complicated too, we must admit that. And the story is not over here, because on top of the above fine structure correction, which is of order α^2, we have afterwards the Lamb shift, which is an order α^3 correction, then the hyperfine splitting, and more.

As usual, we refer to Feynman [35], Griffiths [43], Weinberg [93] for more on all this. In what concerns us, we will be back to such questions in chapters 7-8 below, directly at the advanced level, following Feynman and others, who managed to find a global way of viewing all the phenomena that can appear, corresponding to an infinite series in α.

To be more precise, that theory, called quantum electrodynamics (QED), is an advanced version of quantum mechanics, still used nowadays for any delicate computation. Ironically, while providing an exact answer for the hydrogen atom, QED messes up things too, because that exact answer is not exactly computable. More on this later.

More in detail now, and getting to technicalities, let us review the fine structure of the hydrogen atom, with a proof for the following result, already announced above:

THEOREM 6.7. *The energy levels of the hydrogen atom, taking into account the fine structure coming from the relativistic and spin-related correction, are given by*

$$E_{nj} = E_n \left[1 + \frac{\alpha^2}{n^2} \left(\frac{n}{j+1/2} - \frac{3}{4} \right) \right]$$

with $j = l \pm 1/2$ being as before, and with α being the fine structure constant.

PROOF. We will be doing here something far more precise than what we did before, but still with considerable gaps, namely the usage without proof of some methods from perturbation theory, for which we refer to Griffiths [43], and then some silence of a deep topic, namely the formula of the magnetic dipole of the electron, for which we refer to Griffiths [43] and Weinberg [93]. Anyway, here is what we have, for what's worth:

(1) We will use a general principle from perturbation theory, stating that the perturbed energy appears as expectation of the added Hamiltonian operator:

$$\mathcal{E} = < \mathcal{H} >$$

(2) Let us compute first the relativistic correction. According to the Einstein energy formula, from special relativity, the relativistic kinetic energy is given by:

$$T = \frac{p^2}{2m} - \frac{p^4}{8m^3c^2} + \ldots$$

The Schrödinger equation, which is based on the non-relativistic formula for kinetic energy $T = p^2/2m$, must be therefore corrected with a term as follows:

$$\mathcal{T} = -\frac{p^4}{8m^3c^2}$$

But this leads to the following correction term $\mathcal{E}_{nl}$, with the computation using the Schrödinger equation $p^2\psi = 2m(E-V)\psi$ at the end:

$$\begin{aligned} \mathcal{E}_{nl} &= < \mathcal{T} > \\ &= -\frac{1}{8m^3c^2} < p^4\psi, \psi > \\ &= -\frac{1}{8m^3c^2} < p^2\psi, p^2\psi > \\ &= -\frac{1}{2mc^2} < (E_n - V)^2 > \end{aligned}$$

(3) Now by expanding and using some standard expectation computations, for which we refer as before to Griffiths [43], all in terms of the Bohr radius a, we obtain:

$$\begin{aligned}
\mathcal{E}_{nl} &= -\frac{1}{2mc^2}\left(E_n^2 - 2E_n < V > + < V^2 >\right) \\
&= -\frac{1}{2mc^2}\left(E_n^2 + 2E_nKe^2\left\langle\frac{1}{r}\right\rangle + (Ke^2)^2\left\langle\frac{1}{r^2}\right\rangle\right) \\
&= -\frac{1}{2mc^2}\left(E_n^2 + \frac{2E_nKe^2}{n^2a} + \frac{(Ke^2)^2}{(l+1/2)n^3a^2}\right) \\
&= -\frac{E_n}{2mc^2}\left(\frac{4n}{l+1/2} - 3\right) \\
&= \frac{\alpha^2E_n}{n^2}\left(\frac{n}{l+1/2} - \frac{3}{4}\right)
\end{aligned}$$

(4) Regarding now the spin correction, the reasons for it are very intuitive. The electron, as any spinning charge, has a magnetic field, and the proton, which moves around the electron, will extert some torque on this magnetic field, which must be taken into account, and so which will modify the Hamiltonian, and also the energy E_n.

(5) In practice now, in order to compute the correction we will use the same method as before, namely perturbation theory, but the data will be more delicate to gather.

(6) To start with, when regarding the electron as being fixed, the proton turns around it, and formes an electric loop. The intensity of the corresponding magnetic field can be computed by using the Biot-Savart law, the formula being as follows, with $I = e/T$ being the intensity of the current, and T being the loop time:

$$\begin{aligned}
||B|| &= \frac{\mu_0 I}{2r} \\
&= \frac{\mu_0 e}{2rT} \\
&= \frac{2K\pi e}{c^2rT}
\end{aligned}$$

On the other hand, recall that the angular momentum of the electron satisfies:

$$||L|| = rm||v|| = \frac{2\pi mr^2}{T}$$

Since both B and L point in the same direction, we conclude that we have:

$$B = \frac{Ke}{mc^2r^3}\,L$$

(7) The thing now is that B acts on the magnetic dipole of the electron, which is formally given, for our computation, by the following formula:

$$M=-\frac{e}{2m}\cdot S$$

And there is a long story with this formula, because this is what comes out from a classical electrodynamics computation, so things looking fine. But there is a correction to be made to it, consisting of a rather standard $1/2$ factor called Thomas precession, and on the other hand there is another correction to it, found by Dirac via a non-trivial relativistic computation, consisting of a 2 factor. And these $1/2$ and 2 factors kill each other. For more on the story here, we refer to Griffiths [43] or Weinberg [93].

(8) Moving ahead, based on the formulae of B,M found above, we can compute the correction to the Hamiltonian operator to be made, which is given by:

$$\mathcal{H}=<B,M>=\frac{Ke^2}{2}\cdot\frac{1}{m^2c^2r^3}<L,S>$$

Thus, we are now in familiar territory, and we can use perturbation theory. By skipping some details here, the correction to the energy formula is as follows, with $J=L+S$:

$$\begin{aligned}
\mathcal{E}_{nj}&=<\mathcal{H}>\\
&=\frac{Ke^2}{4}\cdot\frac{1}{m^2c^2r^3}<J^2-L^2-S^2>\\
&=\frac{Ke^2}{4}\cdot\frac{1}{m^2c^2r^3}\cdot\frac{h^2}{2}\left(j(j+1)-l(l+1)-s(s+1)\right)\left\langle\frac{1}{r^3}\right\rangle\\
&=\frac{Ke^2}{4}\cdot\frac{1}{m^2c^2r^3}\cdot\frac{h^2}{2}\left(j(j+1)-l(l+1)-\frac{3}{4}\right)\cdot\frac{1}{l(l+1/2)(l+1)n^3a^3}\\
&=\frac{E_n^2}{mc^2}\cdot\frac{n\left(j(j+1)-l(l+1)-3/4\right)}{l(l+1/2)(l+1)}\\
&=-\frac{\alpha^2E_n}{n^2}\cdot\frac{n(j-l)}{(l+1/2)(j+1/2)}
\end{aligned}$$

(9) Finally, as computed before, the revised energy $E_{nj}=E_n+\mathcal{E}_{nl}+\mathcal{E}_{nj}$, using the formulae in (3,8), is given by the formula in the statement. □

As mentioned before, the story is not over with the above result, because there are several other corrections, which are smaller, coming of top of the fine structure correction, such as the Lamb shift, the hyperfine splitting, and more. In fact, there is an infinite series of corrections, with α as parameter, and the theory designed for solving this problem is quantum electrodynamics (QED), that we will discuss in chapters 7-8 below.

6b.

6c.

6d.

6e. Exercises

Exercises:

Exercise 6.8.

Exercise 6.9.

Exercise 6.10.

Exercise 6.11.

Exercise 6.12.

Exercise 6.13.

Exercise 6.14.

Exercise 6.15.

Bonus exercise.

CHAPTER 7

Dirac equation

7a. Dirac equation

We have seen in previous chapters that quantum mechanics provides explanations and equations for all the basic phenomena appearing at the atomic level. Among others, we have reached to a quite decent level of understanding of the hydrogen atom.

In the remainder of this Part II of the present book we will be interested in rather abstract aspects, and more specifically in "fixing quantum mechanics". And by this we mean not that quantum mechanics is wrong, but that there are certainly a few things that we came upon, which are not very clear, and need to be fixed, as follows:

(1) We would like our theory to be relativistic. Among others, for getting rid of the "relativistic correction" to the hydrogen atom, a correction never being a good thing.

(2) In fact, we would like to have a conceptual understanding of the spin correction too, as to get rid of the whole "fine structure correction" to the hydrogen atom.

(3) We would like our electrons to be joined by more particles, with the minimum here including the protons, the neutrons, and also the photons, representing light.

(4) And then, why not looking too into phenomena that we have not investigated yet, such as radioactivity. Or splitting protons and neutrons into smaller particles.

We will discuss here all these questions. Quite remarkably, there is a common mathematical framework for investigating all these questions, called quantum field theory (QFT). So, we will develop QFT, and then we will turn to questions (1,2) above, and present an amazing answer to them, involving a QFT called quantum electrodynamics (QED). Then we will turn also to questions (3,4), and discuss a bit the status here, notably with a few words on quantum chromodynamics (QCD), which is the quantum field theory obtained by splitting protons and neutrons into smaller particles.

Before starting, let us mention that things won't be easy. Our present level in quantum mechanics, now at this page 150 of the present book, corresponds more or less to things

known since the 1920s. We will of course make big efforts for understanding what happened in the 1930s, then 1940s, then 1950s and so on, but so many things that happened, and the remainder to this Part II will be just a modest introduction to all this.

Getting started, let us formulate a clear objective:

OBJECTIVE 7.1. *We would like to have a relativistic version of quantum mechanics, and with the electron being joined by the photon, representing light. If possible, we would like our theory to cover as well the proton, and the neutron.*

Here the relativistic requirement is very natural in regards with all that has been said above, this being certainly the gate towards a better quantum mechanics.

Regarding the other particles, intuition and common sense would dictate to go first towards the proton and neutron, because aren't these, along with the electron, the constituents of normal matter, that we are normally interested in. However, and here comes our point, mathematically speaking, the electron can certainly live without protons and neutrons, because in order to move, it just needs a positive charge attracting it, and this positive charge can be well something abstract, as per general field theory philosophy.

In contrast, however, the electron cannot live without the photon. The point is that in the context of the basic physics of atoms, electrons can jump between energy levels, emitting or absorbing photons, and with this being known to happen even in the absence of external stimuli. Thus, and for concluding, the true "brother" of the electron is not the proton or the neutron, but rather the photon. And so, the minimal extension of quantum mechanics that we are trying to build should deal with electrons and photons.

Let us first look into the photon, try to understand how to make it fit into our theory, and leave the electron for later. As a starting point, we have:

FACT 7.2. *The master equation for free electromagnetic radiation, that is, for free photons, is the wave equation at speed $v = c$, namely:*

$$\ddot{\varphi} = c^2 \Delta \varphi$$

This equation can be reformulated in the more symmetric form

$$\left(\frac{1}{c^2} \cdot \frac{d^2}{dt^2} - \Delta \right) \varphi = 0$$

with the operator on the left being called the d'Alembertian.

To be more precise here, these are things that we know well, from chapter 5, or even from chapter 2, when first talking about the wave equation, and radiation. In addition, and importantly, we also know from there that the wave equation, at any speed v, is

relativistic, in the sense that it is invariant under Lorentz transformations, which are as follows, with $\gamma = 1/\sqrt{1 - v^2/c^2}$ being as usual the Lorentz factor:

$$x' = \gamma(x - vt)$$
$$y' = y$$
$$z' = z$$
$$t' = \gamma(t - vx/c^2)$$

So far, so good. In relation now with the electron, there is an obvious similarity here with the free Schrödinger equation, without potential V, which reads:

$$\left(i\frac{d}{dt} + \frac{h}{2m}\,\Delta\right)\psi = 0$$

This similarity suggests looking for a relativistic version of the Schrödinger equation, which is compatible with the wave equation at $v = c$. And coming up with such an equation is not very complicated, the straightforward answer being as follows:

DEFINITION 7.3. *The following abstract mathematical equation,*

$$\left(-\frac{1}{c^2}\cdot\frac{d^2}{dt^2} + \Delta\right)\psi = \frac{m^2c^2}{h^2}\,\psi$$

on a function $\psi = \psi_t(x)$, is called the Klein-Gordon equation.

To be more precise, what we have here is some sort of a speculatory equation, formally obtained from the Schrödinger equation, via a few simple manipulations, as to make it relativistic. And with the relation with photons being something very simple, the thing being that at zero mass, $m = 0$, we obtain precisely the wave equation at $v = c$.

All this is very nice, looks like we have a beginning of theory here, both making the electrons relativistic, and unifying them with photons. And isn't this too beautiful to be true. Going ahead now with physics, the following question appears:

QUESTION 7.4. *What does the Klein-Gordon equation really describe?*

And here, unfortunately, bad news all the way. A closer look at the Klein-Gordon equation reveals all sorts of bugs, making it unusable for anything reasonable. And with the main bug, which is enough for disqualifying it, being that, unlike the Schrödinger equation which preserves probability amplitudes $|\psi|^2$, the Klein-Gordon equation does not have this property. Thus, even before trying to understand what the Klein-Gordon equation really describes, we are left with the conclusion that this equation cannot really describe anything reasonable, due to the formal nature of the function ψ involved.

So, this was for the story of the Klein-Gordon equation. Actually this equation was first discovered by Schrödinger himself, in the context of his original work on the Schrödinger equation. But noticing the above bugs with it, Schrödinger dismissed it right

way, and then downgraded his objectives, looking for something non-relativistic instead, and then found the Schrödinger equation, leading to the story that we know.

This being said, the Klein-Gordon equation found later a number of interesting applications, the continuation of the story being as follows:

(1) Dirac found a clever way of extracting the "square root" of the Klein-Gordon equation. And this square root equation, called Dirac equation, turned out to be the correct one, making exactly what the Klein-Gordon equation was supposed to do.

(2) Technically speaking, the Klein-Gordon equation is very useful for investigating the Dirac equation, because the components of the solutions of the Dirac equation satisfy the Klein-Gordon equation. More on this later, when discussing the Dirac equation.

(3) Finally, the Klein-Gordon equation was later recognized to describe well the spin 0 particles. But with these particles being something specialized, including the unstable and sowewhat fringe "pions", and the Higgs boson, which is something complicated.

We will discuss all this, in what follows. In any case, we have here a beginning of good discussion, with our cocktail of thoughts and ideas including electrons, photons, relativity and spin, which are exactly the things that we wanted to include in our discussion. So, all that is left is to clarify all this, and we will do so, following Dirac. Dirac came upon the idea of extracting the square root of the Klein-Gordon operator, as follows:

PROPOSITION 7.5. *We can extract the square root of the Klein-Gordon operator, via a formula as follows,*

$$-\frac{1}{c^2}\cdot\frac{d^2}{dt^2}+\Delta=\left(\frac{i}{c}\cdot\frac{Pd}{dt}+\frac{Qd}{dx}+\frac{Rd}{dy}+\frac{Sd}{dz}\right)^2$$

by using matrices P,Q,R,S which anticommute, $AB=-BA$, and whose squares equal one, $A^2=1$.

PROOF. We have the following computation, valid for any matrices P,Q,R,S, with the notation $\{A,B\}=AB+BA$:

$$\begin{aligned}\left(\frac{i}{c}\cdot\frac{Pd}{dt}+\frac{Qd}{dx}+\frac{Rd}{dy}+\frac{Sd}{dz}\right)^2 &= -\frac{1}{c^2}\cdot\frac{P^2d^2}{dt^2}+\frac{Q^2d^2}{dx^2}+\frac{R^2d^2}{dy^2}+\frac{S^2d^2}{dz^2}\\ &+\frac{i}{c}\left(\frac{\{P,Q\}d^2}{dtdx}+\frac{\{P,R\}d^2}{dtdy}+\frac{\{P,S\}d^2}{dtdz}\right)\\ &+\frac{\{Q,R\}d^2}{dxdy}+\frac{\{Q,S\}d^2}{dxdz}+\frac{\{R,S\}d^2}{dydz}\end{aligned}$$

Thus, in order to obtain in this way the Klein-Gordon operator, the conditions in the statement must be satisfied. □

As a technical comment here, normally when extracting a square root, we should look for a hermitian operator. In view of this, observe that we have:

$$\left(\frac{i}{c}\cdot\frac{Pd}{dt}+\frac{Qd}{dx}+\frac{Rd}{dy}+\frac{Sd}{dz}\right)^*=-\frac{i}{c}\cdot\frac{P^*d}{dt}+\frac{Q^*d}{dx}+\frac{R^*d}{dy}+\frac{S^*d}{dz}$$

Thus, we should normally add the conditions $P^*=-P$ and $Q^*=Q$, $R^*=R$, $S^*=S$ to those above. But, the thing is that due to some subtle reasons, the natural square root of the Klein-Gordon operator is not hermitian. More on this later.

Looking for matrices P,Q,R,S as above is not exactly trivial, and the simplest solutions appear in $M_4(\mathbb{C})$, in connection with the Pauli matrices, as follows:

PROPOSITION 7.6. *The simplest matrices P,Q,R,S as above appear as*

$$P=\gamma_0\quad,\quad Q=i\gamma_1\quad,\quad R=i\gamma_2\quad,\quad S=i\gamma_3$$

with $\gamma_0,\gamma_1,\gamma_2,\gamma_3$ being the Dirac matrices, given by

$$\gamma_0=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\quad,\quad\gamma_i=\begin{pmatrix}0&\sigma_i\\-\sigma_i&0\end{pmatrix}$$

where $\sigma_1,\sigma_2,\sigma_3$ are the Pauli spin matrices, given by:

$$\sigma_1=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\quad,\quad\sigma_2=\begin{pmatrix}0&1\\1&0\end{pmatrix}\quad,\quad\sigma_3=\begin{pmatrix}0&-i\\i&0\end{pmatrix}$$

PROOF. We have $\gamma_0^2=1$, and by using $\sigma_i^2=1$ for any $i=1,2,3$, we have as well the following formula, which shows that we have $(i\gamma_i)^2=1$, as needed:

$$\gamma_i^2=\begin{pmatrix}0&\sigma_i\\-\sigma_i&0\end{pmatrix}\begin{pmatrix}0&\sigma_i\\-\sigma_i&0\end{pmatrix}=\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$$

As in what regards the commutators, we first have, for any $i=1,2,3$, the following equalities, which show that γ_0 anticommutes indeed with γ_i:

$$\gamma_0\gamma_i=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\begin{pmatrix}0&\sigma_i\\-\sigma_i&0\end{pmatrix}=\begin{pmatrix}0&\sigma_i\\\sigma_i&0\end{pmatrix}$$

$$\gamma_i\gamma_0=\begin{pmatrix}0&\sigma_i\\-\sigma_i&0\end{pmatrix}\begin{pmatrix}1&0\\0&-1\end{pmatrix}=\begin{pmatrix}0&-\sigma_i\\-\sigma_i&0\end{pmatrix}$$

Regarding now the remaining commutators, observe here that we have:

$$\gamma_i\gamma_j=\begin{pmatrix}0&\sigma_i\\-\sigma_i&0\end{pmatrix}\begin{pmatrix}0&\sigma_j\\-\sigma_j&0\end{pmatrix}=\begin{pmatrix}-\sigma_i\sigma_j&0\\0&-\sigma_i\sigma_j\end{pmatrix}$$

Now since the Pauli matrices anticommute, we obtain $\gamma_i\gamma_j=-\gamma_j\gamma_i$, as desired. □

We can now put everything together, and we obtain:

THEOREM 7.7. *The following operator, called Dirac operator,*

$$D = i\left(\frac{\gamma_0 d}{cdt} + \frac{\gamma_1 d}{dx} + \frac{\gamma_2 d}{dy} + \frac{\gamma_3 d}{dz}\right)$$

has the property that its square is the Klein-Gordon operator.

PROOF. With notations from Proposition 7.5 and Proposition 7.6, and by making the choices in Proposition 7.6, we have:

$$\begin{aligned}\frac{i}{c}\cdot\frac{Pd}{dt} + \frac{Qd}{dx} + \frac{Rd}{dy} + \frac{Sd}{dz} &= \frac{i}{c}\cdot\frac{\gamma_0 d}{dt} + \frac{i\gamma_1 d}{dx} + \frac{i\gamma_2 d}{dy} + \frac{i\gamma_3 d}{dz}\\ &= i\left(\frac{\gamma_0 d}{cdt} + \frac{\gamma_1 d}{dx} + \frac{\gamma_2 d}{dy} + \frac{\gamma_3 d}{dz}\right)\end{aligned}$$

Thus, we have here a square root of the Klein-Gordon operator, as desired. □

We can now extract the square root of the Klein-Gordon equation, as follows:

DEFINITION 7.8. *We have the following equation, called Dirac equation,*

$$ih\left(\frac{\gamma_0 d}{cdt} + \frac{\gamma_1 d}{dx} + \frac{\gamma_2 d}{dy} + \frac{\gamma_3 d}{dz}\right)\psi = mc\psi$$

obtained by extracting the square root of the Klein-Gordon equation.

As usual with such theoretical physics equations, extreme caution is recommended, at least to start with. We will slowly examine this equation, in what follows, and the good news will be that, passed a few difficulties, this will turn to be a true, magic equation.

As a first observation, all this is very related to spin. In fact, as we will see later, the Dirac equation is the correct relativistic equation describing the spin 1/2 particles.

The Dirac equation comes with a price to pay, which is that of opening Pandora's box of particles. To be more precise, once we adopt this equation, we must surely adopt all its free solutions. And bad news here, the solution which is complementary to the electron is not the proton, but rather a weird new particle, called the positron. In order to explain all this, which is something quite tricky, let us start with the following observation:

PROPOSITION 7.9. *For a particle at rest, meaning under the assumption*

$$\frac{d\psi}{dx} = \frac{d\psi}{dy} = \frac{d\psi}{dz} = 0$$

the Dirac equation takes the form

$$\frac{ih}{c}\cdot\gamma_0\cdot\frac{d\psi}{dt} = mc\psi$$

with $\gamma_0 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$ *being as usual the first Dirac matrix.*

PROOF. Consider indeed the Dirac equation, as formulated in Definition 7.8:

$$ih\left(\frac{\gamma_0 d}{cdt}+\frac{\gamma_1 d}{dx}+\frac{\gamma_2 d}{dy}+\frac{\gamma_3 d}{dz}\right)\psi=mc\psi$$

With the above rest assumption, we are led to the equation in the statement. □

The above equation at rest is very easy to solve, the result being as follows:

THEOREM 7.10. *The solutions of the Dirac equation for particles at rest are*

$$\psi=\begin{pmatrix}e^{-imc^2t/h}\xi\\ e^{imc^2t/h}\eta\end{pmatrix}$$

with $\xi,\eta\in\mathbb{R}^2$ being arbitrary vectors.

PROOF. In order to solve the Dirac equation in Proposition 7.9, let us write:

$$\psi=\begin{pmatrix}\varphi\\ \phi\end{pmatrix}$$

With this notation, the Dirac equation at rest takes the following form:

$$\begin{pmatrix}1&0\\0&-1\end{pmatrix}\begin{pmatrix}d\varphi/dt\\ d\phi/dt\end{pmatrix}=-\frac{imc^2}{h}\begin{pmatrix}\varphi\\ \phi\end{pmatrix}$$

Now by looking at the components, the equations are as follows:

$$\frac{d\varphi}{dt}=-\frac{imc^2}{h}\,\varphi\quad,\quad\frac{d\phi}{dt}=\frac{imc^2}{h}\,\phi$$

But the solutions of these latter equations are as follows, with $\xi,\eta\in\mathbb{R}^2$:

$$\varphi=e^{-imc^2t/h}\xi\quad,\quad\phi=e^{imc^2t/h}\eta$$

Thus, we are led to the conclusion in the statement. □

The question is now, is the above result good news or not? Not really, because in view of what we know from quantum mechanics, an $e^{-iEt/h}$ factor should correspond to the time dependence of a quantum state with energy E, which at rest is $E=mc^2$. And from this perspective, while the above φ functions look very good, the other components, the ϕ functions, look bad, seemingly coming from particles having "negative energy".

So, what to do? In order to avoid particles with negative energy, which is something that definitely looks very bad, the solution is that of talking about antiparticles with positive energy, and to formulate, as a continuation of Theorem 7.10:

THEOREM 7.11. *The basic solutions of the Dirac equation for particles at rest are*

$$\psi^1 = e^{-imc^2t/h}\begin{pmatrix}1\\0\\0\\0\end{pmatrix} \quad , \quad \psi^2 = e^{-imc^2t/h}\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$$

corresponding to the electron with spin up, and spin down, plus

$$\psi^3 = e^{imc^2t/h}\begin{pmatrix}0\\0\\1\\0\end{pmatrix} \quad , \quad \psi^4 = e^{imc^2t/h}\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

corresponding to a new particle, the positron, with spin up, and spin down.

PROOF. Here the mathematics comes from what we found in Theorem 7.10, and the terminology and philosophy comes from the above discussion. With the remark that the newly introduced positron is rather an antiparticle, but more on this later. □

Not very good, all this. Dirac himself could not believe it, and it took some joint effort of Weyl, Pauli, Oppenheimer and others to convince him that yes, unfortunately the positrons predicted by his equation are not the usual protons. And so that goodbye reasonable physics, goodbye common sense, and welcome positrons.

In what concerns us, we have been extremely reluctant, throughout this book, to talk about new particles, but no choice now, we will have to back up, and adopt the positrons. But, passed this, we will slam down the cover of Pandora's box, right away. We definitely don't want all sorts of fringe, short-lived particles to invade our theory, and multiply like mushrooms, and transform our carefully built theory into something apocalyptic.

Be said in passing, after some thinking, positrons are not that bad, as particles. If there's one sort of bad particles in this life, these are the short-lived ones, which appear as some sort of "mathematical complications", which do not really exist in the real, statistical life, which takes place over substantial time $t > 0$. And positrons are not like this, they are nice and stable, exactly as the electrons. Their only fault is that of not being very frequent, a positron's fate being that of being quickly eaten by an electron passing by. But to be blamed for this lack of symmetry is not quantum mechanics, but rather the mechanism of the Big Bang, and once we're fine with this, we're fine with positrons.

More about positrons later, when talking about Feynman diagrams and QED. We will see at that time that positrons are in fact something very natural, and we will get to know and love them on the same level as the usual electrons.

Moving now forward, let us attempt to solve the following question:

QUESTION 7.12. *What are the plane wave solutions*

$$\psi(s) = ae^{-i<k,s>}u$$

of the Dirac equation?

To be more precise, we are using here, as argument of the function ψ, the standard relativistic space-time position $s \in \mathbb{R}^4$ of our particle, namely:

$$s = \begin{pmatrix} ct \\ r \end{pmatrix} \quad , \quad r = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Next, we have in the above a constant $a \in \mathbb{R}$, which will be quite irrelevant to our computations, the Dirac equation being linear. Regarding now k, it is convenient to write this vector split over components, as we did in the above with s, as follows:

$$k = \begin{pmatrix} f \\ g \end{pmatrix} \quad , \quad g = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}$$

With these conventions, along with the standard relativistic convention that the space coordinates contribute with $-$ signs, the scalar product in Question 7.12 is given by:

$$< k, s >= cft - < g, r >$$

Now observe that the real part of the exponential in Question 7.12 is given by:

$$Re(e^{-i<k,s>}) = \cos\big(cft - < g, r >\big)$$

Thus, what we have here, justifying the terminology, is a sinusoidal wave propagating in the direction g, with angular frequency and wavelenght as follows:

$$\omega = cf \quad , \quad \lambda = 2\pi/||g||$$

In order to answer Question 7.12, we must first plug into the Dirac equation our special function ψ. We are led in this way to a quite simple equation, as follows:

PROPOSITION 7.13. *The Dirac equation for plane wave functions*

$$\psi(s) = ae^{-i<k,s>}u$$

takes the following special form, no longer involving derivatives,

$$h\big(\gamma_0 f - \gamma_1 g_1 - \gamma_2 g_2 - \gamma_3 g_3\big)u = mcu$$

with the above conventions for indices and vectors.

PROOF. Consider indeed the Dirac equation, as formulated in Definition 7.8:

$$ih\left(\frac{\gamma_0 d}{cdt} + \frac{\gamma_1 d}{dx} + \frac{\gamma_2 d}{dy} + \frac{\gamma_3 d}{dz}\right)\psi = mc\psi$$

For the function ψ in the statement, the derivatives are given by:

$$\frac{d\psi}{ds_i} = -ik_i\psi$$

Thus, with our above conventions for indices and vectors, we have:

$$\frac{d\psi}{cdt} = -if\psi \quad , \quad \frac{d\psi}{dr_i} = ig_i\psi$$

By plugging these quantities in the Dirac equation, this equation becomes:

$$h\big(\gamma_0 f - \gamma_1 g_1 - \gamma_2 g_2 - \gamma_3 g_3\big)\psi = mc\psi$$

Now by using again $\psi = ae^{-i<k,s>}u$, this equation takes the following form:

$$h\big(\gamma_0 f - \gamma_1 g_1 - \gamma_2 g_2 - \gamma_3 g_3\big)ae^{-i<k,s>}u = mcae^{-i<k,s>}u$$

Thus, by simplifying, we are led to the equation in the statement. □

Let us study now the equation that we found. As a first observation, we can further fine-tune the equation in Proposition 7.13, via some simple manipulations, as follows:

PROPOSITION 7.14. *In the context of Proposition 7.13, with the notation*

$$u = \begin{pmatrix} v \\ w \end{pmatrix}$$

the Dirac equation takes the following form, in terms of the components v, w,

$$v = \frac{< g, \sigma >}{f - mc/h}\, w \qquad , \qquad w = \frac{< g, \sigma >}{f + mc/h}\, v$$

where $\sigma_1, \sigma_2, \sigma_3$ stand as usual for the Pauli spin matrices.

PROOF. According to the definition of the Dirac matrices, in terms of the Pauli ones, we have the following computation, for the operator appearing in Proposition 7.13:

$$\begin{aligned} \gamma_0 f - \gamma_1 g_1 - \gamma_2 g_2 - \gamma_3 g_3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} f - \sum_{i=1}^{3} \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix} g_i \\ &= \begin{pmatrix} f & 0 \\ 0 & -f \end{pmatrix} - \begin{pmatrix} 0 & < g, \sigma > \\ - < g, \sigma > & 0 \end{pmatrix} \\ &= \begin{pmatrix} f & - < g, \sigma > \\ < g, \sigma > & -f \end{pmatrix} \end{aligned}$$

Thus, the quantity which must vanish in Proposition 7.13 is given by:

$$\begin{aligned}
&\Big(h\big(\gamma_0 f-\gamma_1 g_1-\gamma_2 g_2-\gamma_3 g_3\big)-mc\Big)u\\
=\ &\begin{pmatrix} hf-mc & -h<g,\sigma> \\ h<g,\sigma> & -hf-mc\end{pmatrix}\begin{pmatrix} v\\ w\end{pmatrix}\\
=\ &\begin{pmatrix} (hf-mc)v-h<g,\sigma>w\\ h<g,\sigma>v-(hf+mc)w\end{pmatrix}
\end{aligned}$$

We therefore conclude that, in our case, the Dirac equation reads:

$$(hf-mc)v=h<g,\sigma>w$$

$$h<g,\sigma>v=(hf+mc)w$$

Thus, we are led to the conclusion in the statement. □

In order to solve now our equation, let us make the following observation:

PROPOSITION 7.15. *In the context of Proposition 7.14 we must have*

$$||g||^2=f^2-\Big(\frac{mc}{h}\Big)^2$$

under the assumption that the solution is nonzero, $u\neq 0$.

PROOF. Consider the equations found in Proposition 7.14, namely:

$$v=\frac{<g,\sigma>}{f-mc/h}\,w\qquad,\qquad w=\frac{<g,\sigma>}{f+mc/h}\,v$$

By substituting, we are led to the following formulae:

$$v=\frac{<g,\sigma>^2}{f^2-(mc/h)^2}\,v\quad,\quad w=\frac{<g,\sigma>^2}{f^2-(mc/h)^2}\,w$$

Thus, assuming that the solution is nonzero, $u\neq 0$, we must have:

$$\frac{<g,\sigma>^2}{f^2-(mc/h)^2}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

Now, let us compute the left term. According to our various conventions above, and to the formulae for the Pauli matrices, we have the following formula:

$$\begin{aligned}
<g,\sigma>\ &=\ g_1\begin{pmatrix}0&1\\1&0\end{pmatrix}+g_2\begin{pmatrix}0&-i\\i&0\end{pmatrix}+g_3\begin{pmatrix}1&0\\0&-1\end{pmatrix}\\
&=\ \begin{pmatrix}g_3&g_1-ig_2\\g_1+ig_2&-g_3\end{pmatrix}
\end{aligned}$$

By raising this quantity to the square, we obtain:

$$\begin{aligned}
< g, \sigma >^2 &= \begin{pmatrix} g_3 & g_1 - ig_2 \\ g_1 + ig_2 & -g_3 \end{pmatrix} \begin{pmatrix} g_3 & g_1 - ig_2 \\ g_1 + ig_2 & -g_3 \end{pmatrix} \\
&= \begin{pmatrix} g_3^2 + (g_1 - ig_2)(g_1 + ig_2) & g_3(g_1 - ig_2) - (g_1 - ig_2)g_3 \\ (g_1 + ig_2)g_3 - g_3(g_1 + ig_2) & (g_1 + ig_2)(g_1 - ig_2) + g_3^2 \end{pmatrix} \\
&= \begin{pmatrix} g_1^2 + g_2^2 + g_3^2 & 0 \\ 0 & g_1^2 + g_2^2 + g_3^2 \end{pmatrix} \\
&= ||g||^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

Thus, the condition that we found above, coming from $u \neq 0$, reads:

$$\frac{||g||^2}{f^2 - (mc/h)^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

We conclude that we must have the following equality:

$$||g||^2 = f^2 - \left(\frac{mc}{h}\right)^2$$

Thus, we are led to the conclusion in the statement. □

The point now is that the above result invites us to use the rescaled energy-momentum four-vector as variable, $k = \pm p/h$, and we are led in this way to the following result:

THEOREM 7.16. *The basic plane wave solutions, of type*

$$\psi(s) = ae^{-i<k,s>}u$$

of the Dirac equation, come from the functions

$$u^1 = \frac{1}{E + mc^2} \begin{pmatrix} E + mc^2 \\ 0 \\ cp_z \\ cp_x + icp_y \end{pmatrix} \quad , \quad u^2 = \frac{1}{E + mc^2} \begin{pmatrix} 0 \\ E + mc^2 \\ cp_x - icp_y \\ -cp_z \end{pmatrix}$$

corresponding to particle solutions, plus from the functions

$$u^3 = \frac{1}{E + mc^2} \begin{pmatrix} cp_z \\ cp_x + icp_y \\ E + mc^2 \\ 0 \end{pmatrix} \quad , \quad u^4 = \frac{1}{E + mc^2} \begin{pmatrix} cp_x - icp_y \\ -cp_z \\ 0 \\ E + mc^2 \end{pmatrix}$$

corresponding to antiparticle solutions.

PROOF. This comes by putting together all the above. Indeed, with $k = \pm p/h$, as suggested above, we have four choices, which are as follows:

$$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , \quad w = \frac{c}{E + mc^2} \begin{pmatrix} p_z \\ p_x + ip_y \end{pmatrix}$$

$$v = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad , \quad w = \frac{c}{E + mc^2} \begin{pmatrix} p_x - ip_y \\ -p_z \end{pmatrix}$$

$$w = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , \quad v = \frac{c}{E + mc^2} \begin{pmatrix} p_z \\ p_x + ip_y \end{pmatrix}$$

$$w = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad , \quad v = \frac{c}{E + mc^2} \begin{pmatrix} p_x - ip_y \\ -p_z \end{pmatrix}$$

Thus, we are led to the solutions in the statement. □

Regarding the exact physical interpretation of the above plane wave solutions that we found, this is something quite tricky, and we will discuss this later.

In any case, we have now in our theory the electron accompanied by the positron and the photon. There are in fact many other particles which satisfy the Dirac equation, with this equation being in fact the one which describes the spin 1/2 particles. More on this later, when we will know more about the various particles that can appear.

As a last topic, from this preliminary discussion on the Dirac equation, let us discuss now the normalization of the solutions that we found above. We will need:

PROPOSITION 7.17. *For the basic plane wave solutions found above, we have*

$$||u||^2 = \frac{2E}{E + mc^2}$$

with the norm being computed with respect to the usual complex scalar product.

PROOF. According to our formulae above, for $u = u^1, u^2, u^3, u^4$ we have:

$$\begin{aligned} ||u||^2 &= \frac{1}{(E + mc^2)^2} \left((E + mc^2)^2 + c^2(p_x^2 + p_y^2 + p_z^2) \right) \\ &= \frac{1}{(E + mc^2)^2} \left((E + mc^2)^2 + c^2||p||^2 \right) \end{aligned}$$

Now recall that for the energy-momentum vector $\tilde{p} = (E/c, p)$ we have $||\tilde{p}|| = mc$. Thus, the norm of the momentum vector component is given by:

$$||p||^2 = \left(\frac{E}{c} \right)^2 - ||\tilde{p}||^2 = \frac{E^2}{c^2} - m^2c^2$$

With this formula in hand, we can finish our computation, as follows:

$$\begin{aligned}
||u||^2 &= \frac{1}{(E+mc^2)^2}\left((E+mc^2)^2+c^2\left(\frac{E^2}{c^2}-m^2c^2\right)\right)\\
&= \frac{1}{(E+mc^2)^2}\left(E^2+m^2c^4+2Emc^2+E^2-m^2c^4\right)\\
&= \frac{1}{(E+mc^2)^2}\left(2E^2+2Emc^2\right)\\
&= \frac{2E}{E+mc^2}
\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

In what regards now the normalization of the solutions u found in Theorem 7.16, there are several possible useful conventions here, as follows:

$$||Nu||^2=\frac{2E}{c}\quad,\quad ||Nu||^2=\frac{E}{mc^2}\quad,\quad ||Nu||^2=1$$

The corresponding normalizations constants N can be computed by using Proposition 7.17, and are respectively given by the following formulae:

$$N=\sqrt{\frac{E+mc^2}{c}}\quad,\quad N=\sqrt{\frac{E+mc^2}{2mc^2}}\quad,\quad N=\sqrt{\frac{E+mc^2}{2E}}$$

As before with the exact physical interpretation of the plane wave solutions that we found, their normalization is also something quite tricky, and we will discuss this later.

Let us discuss now invariance questions for the solutions of the Dirac equation. As already mentioned in the above, this equation was meant to be a relativistic version of the Schrödinger equation, but the fact that this equation is indeed relativistic, from the point of view of the invariance of solutions, is still something that we must establish.

We recall that the relativistic frame change, with respect to moving with speed v along Ox, is given by the following formulae, where $\beta=v/c$ and $\gamma=1/\sqrt{1-\beta^2}$:

$$\begin{aligned}
ct' &= \gamma(ct-\beta x)\\
x' &= \gamma(x-\beta ct)\\
y' &= y\\
z' &= z
\end{aligned}$$

Equivalently, in matrix form, we have the following formula:

$$\begin{pmatrix} ct'\\ x'\\ y'\\ z'\end{pmatrix}=\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0\\ -\gamma\beta & \gamma & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix} ct\\ x\\ y\\ z\end{pmatrix}$$

Regarding the reverse frame change, this is obtained via $v \to -v$, which gives the following formulae, with $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$ as before:

$$ct = \gamma(ct' + \beta x')$$

$$x = \gamma(x' + \beta ct')$$

$$y = y'$$

$$z = z'$$

Equivalently, in matrix form, we have the following formula:

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix}$$

We refer to the above for more on these formulae, and also for a proof of the fact that the Maxwell equations are indeed invariant under these transformations.

In what regards now the Dirac equation, we have the following result:

THEOREM 7.18. *A solution ψ of the Dirac equation leads, infinitesimally, to the following solution of the same equation, with respect to a frame change as above,*

$$\psi' = A\psi$$

with the matrix A being given by the following formula,

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

where the parameters are given by the following formulae,

$$a = \sqrt{\frac{\gamma+1}{2}} \quad , \quad b = -\sqrt{\frac{\gamma-1}{2}}$$

with $\gamma = 1/\sqrt{1-v^2/c^2}$ being the Lorentz factor.

PROOF. This is something quite tricky, the idea being as follows:

(1) Consider indeed the Dirac equation, as formulated in Definition 7.8:

$$ih\left(\frac{\gamma_0 d}{cdt} + \frac{\gamma_1 d}{dx} + \frac{\gamma_2 d}{dy} + \frac{\gamma_3 d}{dz}\right)\psi = mc\psi$$

It is convenient to use the relativistic space-time position vector, given by:

$$s = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

With this convention, the Dirac equation, as formulated above, becomes:

$$ih \sum_{i=0}^{3} \gamma_i \frac{d\psi}{ds_i} = mc\psi$$

(2) Now let us write as well this equation in the new frame, as follows:

$$ih \sum_{i=0}^{3} \gamma_i \frac{d\psi'}{ds'_i} = mc\psi'$$

We can compute the derivation operators d/ds'_i in terms of the original derivation operators d/ds_i by using the chain rule, starting from:

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix}$$

Indeed, if we denote by L^{-1} the 4×4 matrix appearing above, that of the reverse frame change, then the above formula reads, in terms of space-time position vectors:

$$s = L^{-1}s'$$

Now by using the chain rule, we obtain from this the following formula:

$$\begin{aligned}\frac{d}{ds'_i} &= \sum_j \frac{ds_j}{ds'_i}\cdot\frac{d}{ds_j}\\ &= \sum_j \frac{d(L^{-1}s')_j}{ds'_i}\cdot\frac{d}{ds_j}\\ &= \sum_{jk} \frac{d((L^{-1})_{jk}s'_k)}{ds'_i}\cdot\frac{d}{ds_j}\\ &= \sum_{jk}(L^{-1})_{jk}\frac{ds'_k}{ds'_i}\cdot\frac{d}{ds_j}\\ &= \sum_j(L^{-1})_{ji}\frac{d}{ds_j}\\ &= \sum_j(L^{-1})_{ij}\frac{d}{ds_j}\end{aligned}$$

Here we have used at the end the fact that L^{-1} is symmetric. In vector notation now, the conclusion is that we have the following formula:

$$\frac{d}{ds'} = L^{-1}\frac{d}{ds}$$

(3) With this formula in hand, let us go back to the Dirac equation in the new frame, and try to find a solution of type $\psi' = A\psi$ for it. Our equation reads:

$$ih\sum_{i=0}^{3}\gamma_i\frac{dA\psi}{ds'_i} = mcA\psi$$

By using the linearity of the derivatives, and then the formula found in (2), the left term of this new Dirac equation is given by the following formula:

$$\begin{aligned}ih\sum_{i=0}^{3}\gamma_i\frac{dA\psi}{ds'_i} &= ih\sum_{i=0}^{3}\gamma_i A\frac{d\psi}{ds'_i}\\ &= ih\sum_{i=0}^{3}\gamma_i AL^{-1}\frac{d\psi}{ds_i}\end{aligned}$$

Summarizing, with $\psi' = A\psi$, our equation takes the following form:

$$ih\sum_{i=0}^{3}\gamma_i AL^{-1}\frac{d\psi}{ds_i} = mcA\psi$$

Equivalently, by multiplying everything by A^{-1}, our equation becomes:

$$ih\sum_{i=0}^{3}A^{-1}\gamma_iAL^{-1}\frac{d\psi}{ds_i}=mc\psi$$

(4) Now let us compare this new equation that we found with the original Dirac equation, from (1), which was as follows:

$$ih\sum_{i=0}^{3}\gamma_i\frac{d\psi}{ds_i}=mc\psi$$

In order to have solutions $\psi'=A\psi$ as above, in a plain, non-infinitesimal sense, the obvious possibility is that when we have the following formulae, for any i:

$$A^{-1}\gamma_iAL^{-1}=\gamma_i$$

Thus, as a conclusion to this discussion, in order to prove our theorem, in a plain formulation, it would be enough to establish the following formulae, for any i:

$$A^{-1}\gamma_iA=\gamma_iL$$

(5) With this done, let us have a look at the matrix A in the statement. That matrix is constructed by using two numbers a,b, which are given by:

$$a=\sqrt{\frac{\gamma+1}{2}}\quad,\quad b=-\sqrt{\frac{\gamma-1}{2}}$$

Our first claim is that we have the following useful formulae, relating a,b:

$$a^2-b^2=1$$
$$a^2+b^2=\gamma$$
$$2ab=-\gamma\beta$$

Indeed, the first two formulae are clear, and the third formula comes from:

$$\begin{aligned}2ab&=-\sqrt{\gamma^2-1}\\&=-\sqrt{\frac{1}{1-\beta^2}-1}\\&=-\sqrt{\frac{\beta^2}{1-\beta^2}}\\&=-\frac{\beta}{\sqrt{1-\beta^2}}\\&=-\gamma\beta\end{aligned}$$

Observe also that the above formula $a^2-b^2=1$ suggests using a notation of type $a=\cosh p,b=\sinh p$, but we will not need this here.

(6) Before getting to the matrix A in the statement, let us further study the above numbers a, b. With the help of the formulae connecting them, from (5), we obtain:

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} a & b \\ b & a \end{pmatrix} = \begin{pmatrix} a^2+b^2 & 2ab \\ 2ab & a^2+b^2 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix}$$

We recognize here the upper left block of L, and so we have:

$$L = \begin{pmatrix} a & b & 0 & 0 \\ b & a & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^2$$

A similar discussion goes for the inverse Lorentz matrix. Indeed, we have:

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} a & -b \\ -b & a \end{pmatrix} = \begin{pmatrix} a^2-b^2 & 0 \\ 0 & a^2-b^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Thus, we have the following matrix inversion formula:

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

We conclude that the inverse of the Lorentz matrix is given by:

$$L^{-1} = \begin{pmatrix} a & -b & 0 & 0 \\ -b & a & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^2$$

(7) Now let us look at the matrix in the statement, namely:

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

This matrix, and its inverse, are then given by the following formulae:

$$A = a + b\gamma_0\gamma_2$$

$$A^{-1} = a - b\gamma_0\gamma_2$$

Indeed, in what regards the formula of A, this comes from:

$$\begin{aligned}\gamma_0\gamma_2 &= \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}0 & \sigma_2\\ -\sigma_2 & 0\end{pmatrix}\\ &= \begin{pmatrix}0 & \sigma_2\\ \sigma_2 & 0\end{pmatrix}\\ &= \begin{pmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\end{pmatrix}\end{aligned}$$

As for the formula of A^{-1}, this comes from the following computation, with $J = \gamma_0\gamma_2$, which satisfies $J^2 = 1$, by using the formula $a^2 - b^2 = 1$ from (5):

$$\begin{aligned}(a + bJ)(a - bJ) &= a^2 + abJ - abJ - b^2J^2\\ &= a^2 - b^2\\ &= 1\end{aligned}$$

(8) In relation now with the formulae needed in (4), our first claim is that:

$$A\gamma_0 A = \gamma_0$$

$$A^{-1}\gamma_1 A = \gamma_1$$

$$A\gamma_2 A = \gamma_2$$

$$A^{-1}\gamma_3 A = \gamma_3$$

(9) Indeed, the first formula comes from the following computation:

$$\begin{aligned}A\gamma_0 A &= (a + b\gamma_0\gamma_2)\gamma_0(a + b\gamma_0\gamma_2)\\ &= a^2\gamma_0 + ab\gamma_0\gamma_0\gamma_2 + ab\gamma_0\gamma_2\gamma_0 + b^2\gamma_0\gamma_2\gamma_0\gamma_0\gamma_2\\ &= a^2\gamma_0 - b^2\gamma_0\\ &= \gamma_0\end{aligned}$$

The second formula comes from a similar computation, as follows:

$$\begin{aligned}A^{-1}\gamma_1 A &= (a + b\gamma_0\gamma_2)\gamma_1(a - b\gamma_0\gamma_2)\\ &= a^2\gamma_1 - ab\gamma_1\gamma_0\gamma_2 + ab\gamma_0\gamma_2\gamma_1 - b^2\gamma_0\gamma_2\gamma_1\gamma_0\gamma_2\\ &= a^2\gamma_1 - b^2\gamma_1\\ &= \gamma_1\end{aligned}$$

The third formula again comes from a similar computation, as follows:

$$\begin{aligned} A\gamma_2 A &= (a+b\gamma_0\gamma_2)\gamma_2(a+b\gamma_0\gamma_2)\\ &= a^2\gamma_2+ab\gamma_2\gamma_0\gamma_2+ab\gamma_0\gamma_2\gamma_2+b^2\gamma_0\gamma_2\gamma_2\gamma_0\gamma_2\\ &= a^2\gamma_2-b^2\gamma_2\\ &= \gamma_2 \end{aligned}$$

As for the fourth formula, this comes again from a similar computation, namely:

$$\begin{aligned} A^{-1}\gamma_3 A &= (a+b\gamma_0\gamma_2)\gamma_3(a-b\gamma_0\gamma_2)\\ &= a^2\gamma_3-ab\gamma_3\gamma_0\gamma_2+ab\gamma_0\gamma_2\gamma_3-b^2\gamma_0\gamma_2\gamma_3\gamma_0\gamma_2\\ &= a^2\gamma_3-b^2\gamma_3\\ &= \gamma_3 \end{aligned}$$

(10) Now observe that, with respect to the formulae needed in (4), the second and the fourth formulae found in (8) are what we need. As for the first and third formulae, these are not exactly what we need, and we must fine-tune them. We first have:

$$\begin{aligned} A^{-1}\gamma_0 A &= (a-b\gamma_0\gamma_2)\gamma_0(a+b\gamma_0\gamma_2)\\ &= a^2\gamma_0+ab\gamma_0\gamma_0\gamma_2-ab\gamma_0\gamma_2\gamma_0-b^2\gamma_0\gamma_2\gamma_0\gamma_0\gamma_2\\ &= (a^2+b^2)\gamma_0+2ab\gamma_2\\ &= \gamma\cdot\gamma_0-\gamma\beta\cdot\gamma_2 \end{aligned}$$

Similarly, we have the following computation:

$$\begin{aligned} A^{-1}\gamma_2 A &= (a-b\gamma_0\gamma_2)\gamma_2(a+b\gamma_0\gamma_2)\\ &= a^2\gamma_2+ab\gamma_2\gamma_0\gamma_2-ab\gamma_0\gamma_2\gamma_2-b^2\gamma_0\gamma_2\gamma_2\gamma_0\gamma_2\\ &= (a^2+b^2)\gamma_2-2ab\gamma_0\\ &= \gamma\cdot\gamma_2+\gamma\beta\cdot\gamma_0 \end{aligned}$$

(11) Time now to review the conditions found in (4). These conditions, corresponding to the plain Lorentz invariance of the solutions of the Dirac equation, were $A^{-1}\gamma_i A=\gamma_i L$. But because of $\gamma_i^2=1$, we can reformulate them in the following way:

$$L=\gamma_i A^{-1}\gamma_i A$$

Now in view of the above, it makes sense to introduce the following matrices:

$$L_i=\gamma_i A^{-1}\gamma_i A$$

According to the computations in (9), we have the following formulae:

$$L_1=L_3=1$$

On the other hand, according to the computations in (10), we have as well:

$$L_0=\gamma_0(\gamma\cdot\gamma_0-\gamma\beta\cdot\gamma_2)=\gamma-\gamma\beta\cdot\gamma_0\gamma_2$$

$$L_2 = \gamma_2(\gamma \cdot \gamma_2 + \gamma\beta \cdot \gamma_0) = \gamma - \gamma\beta \cdot \gamma_0\gamma_2$$

Thus, in usual matrix form, we have the following formulae:

$$L_0 = L_2 = \begin{pmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & \gamma & -\gamma\beta & 0 \\ 0 & -\gamma\beta & \gamma & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{pmatrix}$$

(12) The point now is that, based on what we found above, we can say that $\psi' = A\psi$ satisfies the Dirac equation in the new frame, in an infinitesimal sense, as claimed. □

7b.

7c.

7d.

7e. Exercises

Exercises:

Exercise 7.19.

Exercise 7.20.

Exercise 7.21.

Exercise 7.22.

Exercise 7.23.

Exercise 7.24.

Exercise 7.25.

Exercise 7.26.

Bonus exercise.

CHAPTER 8

Feynman diagrams

Exercises:

Exercise 8.1.

Exercise 8.2.

Exercise 8.3.

Exercise 8.4.

Exercise 8.5.

Exercise 8.6.

Exercise 8.7.

Exercise 8.8.

Bonus exercise.

Part III

Bosons, fermions

Come back home to the refinery
Hiring man says son, if it was up to me
Went down to see my V.A. man
He said son, don't you understand

CHAPTER 9

Particle physics

Exercises:

Exercise 9.1.

Exercise 9.2.

Exercise 9.3.

Exercise 9.4.

Exercise 9.5.

Exercise 9.6.

Exercise 9.7.

Exercise 9.8.

Bonus exercise.

CHAPTER 10

Bosons, fermions

10a. Bosons, fermions

10b.

10c.

10d.

10e. Exercises

Exercises:

Exercise 10.1.

Exercise 10.2.

Exercise 10.3.

Exercise 10.4.

Exercise 10.5.

Exercise 10.6.

Exercise 10.7.

Exercise 10.8.

Bonus exercise.

CHAPTER 11

Statistical mechanics

Exercises:

Exercise 11.1.

Exercise 11.2.

Exercise 11.3.

Exercise 11.4.

Exercise 11.5.

Exercise 11.6.

Exercise 11.7.

Exercise 11.8.

Bonus exercise.

CHAPTER 12

Exotic matter

Exercises:

Exercise 12.1.

Exercise 12.2.

Exercise 12.3.

Exercise 12.4.

Exercise 12.5.

Exercise 12.6.

Exercise 12.7.

Exercise 12.8.

Bonus exercise.

Part IV

Into molecules

Open to everything happy and sad
Seeing the good, meme si tout va si mal
Voir le soleil quand la nuit nous accable
Oh pour un jour croire aux dieux, croire aux fables

CHAPTER 13

Periodic table

Exercises:

Exercise 13.1.

Exercise 13.2.

Exercise 13.3.

Exercise 13.4.

Exercise 13.5.

Exercise 13.6.

Exercise 13.7.

Exercise 13.8.

Bonus exercise.

CHAPTER 14

Ions, isotopes

Exercises:

Exercise 14.1.

Exercise 14.2.

Exercise 14.3.

Exercise 14.4.

Exercise 14.5.

Exercise 14.6.

Exercise 14.7.

Exercise 14.8.

Bonus exercise.

CHAPTER 15

Small molecules

Exercises:

Exercise 15.1.

Exercise 15.2.

Exercise 15.3.

Exercise 15.4.

Exercise 15.5.

Exercise 15.6.

Exercise 15.7.

Exercise 15.8.

Bonus exercise.

CHAPTER 16

Big molecules

Congratulations for having read this book, and no exercises for this final chapter.

Bibliography

[1] A.A. Abrikosov, Fundamentals of the theory of metals, Dover (1988).

[2] A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinski, Methods of quantum field theory in statistical physics, Dover (1963).

[3] G.W. Anderson, A. Guionnet and O. Zeitouni, An introduction to random matrices, Cambridge Univ. Press (2010).

[4] V.I. Arnold, Ordinary differential equations, Springer (1973).

[5] V.I. Arnold, Mathematical methods of classical mechanics, Springer (1974).

[6] V.I. Arnold, Lectures on partial differential equations, Springer (1997).

[7] V.I. Arnold and B.A. Khesin, Topological methods in hydrodynamics, Springer (1998).

[8] N.W. Ashcroft and N.D. Mermin, Solid state physics, Saunders College Publ. (1976).

[9] M.F. Atiyah, The geometry and physics of knots, Cambridge Univ. Press (1990).

[10] T. Banica, Calculus and applications (2024).

[11] T. Banica, Introduction to modern physics (2024).

[12] T. Banica, The story of particle physics (2024).

[13] G.K. Batchelor, An introduction to fluid dynamics, Cambridge Univ. Press. (1967).

[14] R.J. Baxter, Exactly solved models in statistical mechanics, Academic Press (1982).

[15] I. Bengtsson and K. Życzkowski, Geometry of quantum states, Cambridge Univ. Press (2006).

[16] S.M. Carroll, Spacetime and geometry, Cambridge Univ. Press (2004).

[17] P.M. Chaikin and T.C. Lubensky, Principles of condensed matter physics, Cambridge Univ. Press (1995).

[18] V. Chari and A. Pressley, A guide to quantum groups, Cambridge Univ. Press (1994).

[19] A.R. Choudhuri, Astrophysics for physicists, Cambridge Univ. Press (2012).

[20] D.D. Clayton, Principles of stellar evolution and nucleosynthesis, Univ. of Chicago Press (1968).

[21] A. Connes, Noncommutative geometry, Academic Press (1994).

[22] W.N. Cottingham and D.A. Greenwood, An introduction to the standard model of particle physics, Cambridge Univ. Press (2012).

[23] P.A. Davidson, Introduction to magnetohydrodynamics, Cambridge Univ. Press (2001).

[24] P. Di Francesco, P. Mathieu and D. Sénéchal, Conformal field theory, Springer (1996).

[25] P.A.M. Dirac, Principles of quantum mechanics, Oxford Univ. Press (1930).

[26] M.P. do Carmo, Differential geometry of curves and surfaces, Dover (1976).

[27] M.P. do Carmo, Riemannian geometry, Birkhäuser (1992).

[28] S. Dodelson, Modern cosmology, Academic Press (2003).

[29] A. Einstein, Relativity: the special and the general theory, Dover (1916).

[30] L.C. Evans, Partial differential equations, AMS (1998).

[31] L.D. Faddeev and L. A. Takhtajan, Hamiltonian methods in the theory of solitons, Springer (2007).

[32] E. Fermi, Thermodynamics, Dover (1937).

[33] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics I: mainly mechanics, radiation and heat, Caltech (1963).

[34] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics II: mainly electromagnetism and matter, Caltech (1964).

[35] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics III: quantum mechanics, Caltech (1966).

[36] R.P. Feynman and A.R. Hibbs, Quantum mechanics and path integrals, Dover (1965).

[37] A.P. French, Special relativity, Taylor and Francis (1968).

[38] N. Goldenfeld, Lectures on phase transitions and the renormalization group, CRC Press (1992).

[39] H. Goldstein, C. Safko and J. Poole, Classical mechanics, Addison-Wesley (1980).

[40] D.L. Goodstein, States of matter, Dover (1975).

[41] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory, Cambridge Univ. Press (2012).

[42] D.J. Griffiths, Introduction to electrodynamics, Cambridge Univ. Press (1981).

[43] D.J. Griffiths, Introduction to quantum mechanics, Cambridge Univ. Press (1995).

[44] D.J. Griffiths, Introduction to elementary particles, Wiley (2008).

[45] D.J. Griffiths, Revolutions in twentieth-century physics, Cambridge Univ. Press (2012).

[46] J. Harris, Algebraic geometry, Springer (1992).

[47] W.A. Harrison, Solid state theory, Dover (1970).

[48] W.A. Harrison, Electronic structure and the properties of solids, Dover (1980).

[49] K. Huang, Introduction to statistical physics, CRC Press (2001).

[50] K. Huang, Quantum field theory, Wiley (1998).

[51] K. Huang, Quarks, leptons and gauge fields, World Scientific (1982).

[52] K. Huang, Fundamental forces of nature, World Scientific (2007).

[53] C. Itzykson and J.B. Zuber, Quantum field theory, Dover (1980).

[54] J.D. Jackson, Classical electrodynamics, Wiley (1962).

[55] V.F.R. Jones, Subfactors and knots, AMS (1991).

[56] L.P. Kadanoff, Statistical physics: statics, dynamics and renormalization, World Scientific (2000).

[57] T. Kibble and F.H. Berkshire, Classical mechanics, Imperial College Press (1966).

[58] C. Kittel, Introduction to solid state physics, Wiley (1953).

[59] M. Kumar, Quantum: Einstein, Bohr, and the great debate about the nature of reality, Norton (2009).

[60] T. Lancaster and K.M. Blundell, Quantum field theory for the gifted amateur, Oxford Univ. Press (2014).

[61] L.D. Landau and E.M. Lifshitz, Mechanics, Pergamon Press (1960).

[62] L.D. Landau and E.M. Lifshitz, The classical theory of fields, Addison-Wesley (1951).

[63] L.D. Landau and E.M. Lifshitz, Quantum mechanics: non-relativistic theory, Pergamon Press (1959).

[64] V.B. Berestetskii, E.M. Lifshitz and L.P. Pitaevskii, Quantum electrodynamics, Butterworth-Heinemann (1982).

[65] P. Lax, Linear algebra and its applications, Wiley (2007).

[66] P. Lax, Functional analysis, Wiley (2002).

[67] M.A. Nielsen and I.L. Chuang, Quantum computation and quantum information, Cambridge Univ. Press (2000).

[68] W.K.H. Panofsky and M. Phillips, Classical electricity and magnetism, Dover (1955).

[69] R.K. Pathria and and P.D. Beale, Statistical mechanics, Elsevier (1972).

[70] A. Peres, Quantum theory: concepts and methods, Kluwer (1993).

[71] M. Peskin and D.V. Schroeder, An introduction to quantum field theory, CRC Press (1995).

[72] E.M. Purcell and D.J. Morin, Electricity and magnetism, Cambridge Univ. Press (1965).

[73] W. Rudin, Principles of mathematical analysis, McGraw-Hill (1964).

[74] W. Rudin, Real and complex analysis, McGraw-Hill (1966).

[75] B. Ryden, Introduction to cosmology, Cambridge Univ. Press (2002).

[76] B. Ryden and B.M. Peterson, Foundations of astrophysics, Cambridge Univ. Press (2010).

[77] B. Ryden and R.W. Pogge, Interstellar and intergalactic medium, Cambridge Univ. Press (2021).

[78] J.J. Sakurai and J. Napolitano, Modern quantum mechanics, Cambridge Univ. Press (1985).

[79] D.V. Schroeder, An introduction to thermal physics, Oxford Univ. Press (1999).

[80] M. Schwartz, Principles of electrodynamics, Dover (1972).

[81] J. Schwinger, Einstein's legacy: the unity of space and time, Dover (1986).

[82] J. Schwinger, L.L. DeRaad Jr., K.A. Milton and W.Y. Tsai, Classical electrodynamics, CRC Press (1998).

[83] J. Schwinger and B.H. Englert, Quantum mechanics: symbolism of atomic measurements, Springer (2001).

[84] I.R. Shafarevich, Basic algebraic geometry, Springer (1974).

[85] R. Shankar, Fundamentals of physics I: mechanics, relativity, and thermodynamics, Yale Univ. Press (2014).

[86] R. Shankar, Fundamentals of physics II: electromagnetism, optics, and quantum mechanics, Yale Univ. Press (2016).

[87] R. Shankar, Principles of quantum mechanics, Springer (1980).

[88] R. Shankar, Quantum field theory and condensed matter: an introduction, Cambridge Univ. Press (2017).

[89] J. Smit, Introduction to quantum fields on a lattice, Cambridge Univ. Press (2002).

[90] J.R. Taylor, Classical mechanics, Univ. Science Books (2003).

[91] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press (1955).

[92] S. Weinberg, Foundations of modern physics, Cambridge Univ. Press (2011).

[93] S. Weinberg, Lectures on quantum mechanics, Cambridge Univ. Press (2012).

[94] S. Weinberg, Lectures on astrophysics, Cambridge Univ. Press (2019).

[95] S. Weinberg, Cosmology, Oxford Univ. Press (2008).

[96] H. Weyl, The theory of groups and quantum mechanics, Princeton Univ. Press (1931).

[97] H. Weyl, Space, time, matter, Princeton Univ. Press (1918).

[98] J.M. Yeomans, Statistical mechanics of phase transitions, Oxford Univ. Press (1992).

[99] J. Zinn-Justin, Path integrals in quantum mechanics, Oxford Univ. Press (2004).

[100] B. Zwiebach, A first course in string theory, Cambridge Univ. Press (2004).

Index