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. Space and time

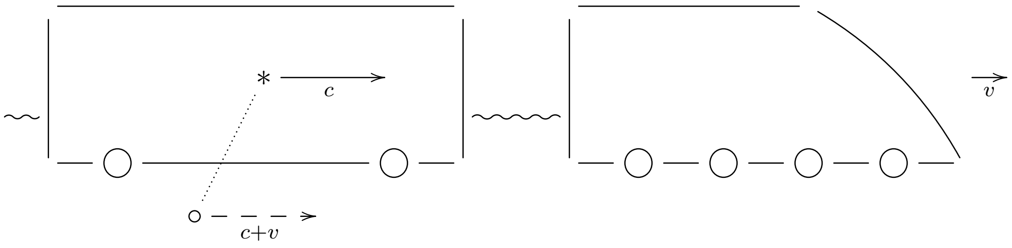
The key to all modern physics, including the quantum mechanics, that we would like to understand in this book, is light. Light is in fact something quite complicated, and more on this later, but as a starting point for our considerations, which is something well-known, and quite intuitive, we have the obvious fact that its speed c is huge.

And, good news, this simple fact, namely that the speed of light c is huge, when compared with all the other speeds that we know, is all we need, for having something started. Indeed, based on experiments by Fizeau, then Michelson-Morley and others, and some physics by Maxwell and Lorentz too, Einstein came upon the following principles:

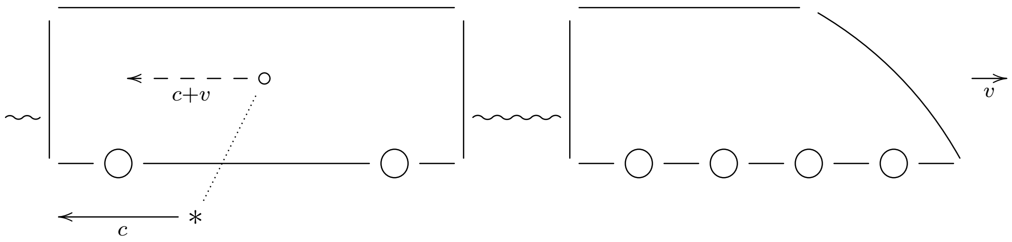
Fact 1.1 (Einstein principles). *The following happen:*

1. *Light travels in vacuum at finite speed, $c < \infty$.*
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, $v \not> c$.*

The point now is that, obviously, something is wrong here. Indeed, assuming for instance that we have a train, running in vacuum at speed $v > 0$, and someone on board lights a flashlight $*$ towards the locomotive, then an observer $\circ$ on the ground will see the light traveling at speed $c + v > c$, which is a contradiction:



Equivalently, with the same train running, in vacuum at speed $v > 0$, if the observer on the ground lights a flashlight $*$ towards the back of the train, then viewed from the train, that light will travel at speed $c+v>c$, which is a contradiction again:



Summarizing, Fact 1.1 implies $c+v=c$, so contradicts classical mechanics, which therefore needs a fix. But the fix in 1D is straightforward, as follows:

Proposition 1.2. *If we sum the speeds according to the Einstein formula*

$$u+_e v=\frac{u+v}{1+uv}$$

in $c=1$ units, then the Galileo formula still holds, approximately, for low speeds

$$u+_e v\simeq u+v$$

and if we have $u=1$ or $v=1$, the resulting sum is $u+_e v=1$.

Proof. All this is self-explanatory, and clear from definitions, and with the Einstein formula of $u+_e v$ itself being just the obvious solution to our $c+v=c$ puzzle. To be more precise, if we plug in $u=1$ in the above summation formula, we obtain as result:

$$1+_e v=\frac{1+v}{1+v}=1$$

And the same happens with $v=1$. Thus, we are led to the above conclusions. $\square$

Let us attempt now to construct $u+_e v$ in arbitrary dimensions. When $u,v\in\mathbb{R}^N$ are proportional, we are basically in 1D, so our addition formula must satisfy:

$$u\sim v\implies u+_e v=\frac{u+v}{1+<u,v>}$$

However, the formula on the right will not work as such in general, for arbitrary speeds $u,v\in\mathbb{R}^N$, and this because we have, as main requirement for our operation, in analogy with the $1+_e v=1$ formula from 1D, the following condition:

$$||u||=1\implies u+_e v=u$$

Equivalently, in analogy with $u+_e 1=1$ from 1D, we would like to have:

$$||v||=1\implies u+_e v=v$$

Summarizing, our $u\sim v$ formula above is not bad, as a start, but we must add a correction term to it, for the above requirements to be satisfied, and of course with the correction term vanishing when $u\sim v$. So, we are led to a math puzzle, as follows:

Puzzle 1.3. *What vanishes when $u \sim v$, and then how to correctly define*

$$u +_e v = \frac{u + v + \gamma_{uv}}{1 + < u, v >}$$

as for the correction term γ_{uv} to vanish when $u \sim v$?

But the solution to the first question is well-known in 3D. Indeed, here we can use the vector product $u \times v$, that we met before, which notoriously satisfies:

$$u \sim v \implies u \times v = 0$$

Thus, our correction term γ_{uv} must be something containing $w = u \times v$, which vanishes when this vector w vanishes, and in addition arranged such that $||u|| = 1$ produces a simplification, with $u +_e v = u$ as end result, and with $||v|| = 1$ producing a simplification too, with $u +_e v = v$ as end result. Thus, our vector calculus puzzle becomes:

Puzzle 1.4. *How to correctly define the Einstein summation in* 3 *dimensions,*

$$u +_e v = \frac{u + v + \gamma_{uvw}}{1 + < u, v >}$$

with $w = u \times v$, in such a way as for the correction term γ_{uvw} to satisfy

$$w = 0 \implies \gamma_{uvw} = 0$$

and also such that $||u|| = 1 \implies u +_e v = u$, and $||v|| \implies u +_e v = v$?

In order to solve this latter puzzle, the first observation is that $\gamma_{uvw} = w$ will not do, and this for several reasons. First, this vector points in the wrong direction, orthogonal to the plane spanned by u, v, and we certainly don't want to leave this plane, with our correction. Also, as a technical remark to be put on top of this, the choice $\gamma_{uvw} = w$ will not bring any simplifications, as required above, in the cases $||u|| = 1$ or $||v|| = 1$. Thus, certainly wrong choice, and we must invent something more complicated.

Moving ahead now, as obvious task, we must "transport" the vector w to the plane spanned by u, v. But this is simplest done by taking the vector product with any vector in this plane, and we are led in this way to an update of our puzzle, as follows:

Puzzle 1.5. *How to define the Einstein summation in* 3 *dimensions,*

$$u +_e v = \frac{u + v + \gamma_{uvw}}{1 + < u, v >}$$

with the correction term being of the following form, with $w = u \times v$, and $\alpha, \beta \in \mathbb{R}$,

$$\gamma_{uvw} = (\alpha u + \beta v) \times w$$

in such a way as to have $||u|| = 1 \implies u +_e v = u$, and $||v|| \implies u +_e v = v$?

In order to investigate what happens when $||u||=1$ or $||v||=1$, we must compute the vector products $u\times w$ and $v\times w$. So, pausing now our study for consulting the vector calculus database, and then coming back, here is the formula that we need:

$$u\times(u\times v)=<u,v>u-<u,u>v$$

As for the formula of $v\times w$, that I forgot to record, we can recover it from the one above of $u\times w$, by using the basic properties of the vector products, as follows:

$$\begin{aligned}v\times(u\times v)&=-v\times(v\times u)\\&=-(<v,u>v-<v,v>u)\\&=<v,v>u-<u,v>v\end{aligned}$$

With these formulae in hand, we can now compute the correction term, with the result here, that we will need several times in what comes next, being as follows:

Proposition 1.6. *The correction term* $\gamma_{uvw}=(\alpha u+\beta v)\times w$ *is given by*

$$\gamma_{uvw}=(\alpha<u,v>+\beta<v,v>)u-(\alpha<u,u>+\beta<u,v>)v$$

for any values of the scalars $\alpha,\beta\in\mathbb{R}$.

Proof. According to our vector product formulae above, we have:

$$\begin{aligned}\gamma_{uvw}&=(\alpha u+\beta v)\times w\\&=\alpha(<u,v>u-<u,u>v)+\beta(<v,v>u-<u,v>v)\\&=(\alpha<u,v>+\beta<v,v>)u-(\alpha<u,u>+\beta<u,v>)v\end{aligned}$$

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

Time now to see what happens when $||u||=1$ and $||v||=1$, if we can get indeed $u+_e v=u$ and $u+_e v=v$. But this is not really possible, so time to breathe, decide that we did enough work for the day, and formulate our conclusions as follows:

Proposition 1.7. *When defining the Einstein speed summation in* 3D *as*

$$u+_e v=\frac{u+v+u\times(u\times v)}{1+<u,v>}$$

in $c=1$ *units, the following happen:*

1. *When* $u\sim v$, *we recover the previous* 1D *formula.*
2. *When* $||u||=1$, *speed of light, we have* $u+_e v=u$.
3. *However,* $||v||=1$ *does not imply* $u+_e v=v$.
4. *Also, the formula* $u+_e v=v+_e u$ *fails.*

Proof. This follows indeed from the above discussion, with the following choice for the correction term, by favoring the $||u||=1$ problem over the $||v||=1$ one:

$$\gamma_{uvw}=u\times w$$

In fact, with this choice made, the computation is very simple, as follows:

$$\begin{aligned} \|u\|=1 \quad &\implies \quad \gamma_{uvw}=<u,v>u-v \\ &\implies \quad u+v+\gamma_{uvw}=u+<u,v>u \\ &\implies \quad \frac{u+v+\gamma_{uvw}}{1+<u,v>}=u \end{aligned}$$

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

With this done, shall we give up? We can in fact do even better, as follows:

Theorem 1.8. *When defining the Einstein speed summation in 3D as*

$$u+_e v=\frac{1}{1+<u,v>}\left(u+v+\frac{u\times(u\times v)}{1+\sqrt{1-\|u\|^2}}\right)$$

in $c=1$ units, the following happen:

1. *When $u\sim v$, we recover the previous 1D formula.*
2. *We have $\|u\|,\|v\|<1\implies\|u+_e v\|<1$.*
3. *When $\|u\|=1$, we have $u+_e v=u$.*
4. *When $\|v\|=1$, we have $\|u+_e v\|=1$.*
5. *However, $\|v\|=1$ does not imply $u+_e v=v$.*
6. *Also, the formula $u+_e v=v+_e u$ fails.*

Proof. This follows from the above discussion, with a bit more work, as follows:

(1) This is something that we know from Proposition 1.7.

(2) In order to simplify notation, let us set $\delta=\sqrt{1-\|u\|^2}$, which is the inverse of the quantity $\gamma=1/\sqrt{1-\|u\|^2}$. With this convention, we have:

$$\begin{aligned} u+_e v &= \frac{1}{1+<u,v>}\left(u+v+\frac{<u,v>u-\|u\|^2v}{1+\delta}\right) \\ &= \frac{(1+\delta+<u,v>)u+(1+\delta-\|u\|^2)v}{(1+<u,v>)(1+\delta)} \end{aligned}$$

Taking now the squared norm and computing gives the following formula:

$$\|u+_e v\|^2=\frac{(1+\delta)^2\|u+v\|^2+(\|u\|^2-2(1+\delta))(\|u\|^2\|v\|^2-<u,v>^2)}{(1+<u,v>)^2(1+\delta)^2}$$

But this formula can be further processed by using $\delta=\sqrt{1-\|u\|^2}$, and by navigating through the various quantities which appear, we obtain, as a final product:

$$\|u+_e v\|^2=\frac{\|u+v\|^2-\|u\|^2\|v\|^2+<u,v>^2}{(1+<u,v>)^2}$$

But this type of formula is exactly what we need, for what we want to do. Indeed, by assuming $||u||,||v||<1$, we have the following estimate:

$$\begin{aligned}||u+_e v||^2<1 &\iff ||u+v||^2-||u||^2||v||^2+<u,v>^2<(1+<u,v>)^2\\ &\iff ||u+v||^2-||u||^2||v||^2<1+2<u,v>\\ &\iff ||u||^2+||v||^2-||u||^2||v||^2<1\\ &\iff (1-||u||^2)(1-||v||^2)>0\end{aligned}$$

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

(3) This is something that we know from Proposition 1.7.

(4) This comes from the squared norm formula established in the proof of (2) above, because when assuming $||v||=1$, we obtain:

$$\begin{aligned}||u+_e v||^2 &= \frac{||u+v||^2-||u||^2+<u,v>^2}{(1+<u,v>)^2}\\ &= \frac{||u||^2+1+2<u,v>-||u||^2+<u,v>^2}{(1+<u,v>)^2}\\ &= \frac{1+2<u,v>+<u,v>^2}{(1+<u,v>)^2}\\ &= 1\end{aligned}$$

(5) This is clear, from the obvious lack of symmetry of our formula.

(6) This is again clear, from the obvious lack of symmetry of our formula. □

That was nice, all this mathematics, and hope you're still with me. And good news, the formula in Theorem 1.8 is the good one, confirmed by experimental physics.

We can further build on the above, with the following surprising conclusions:

THEOREM 1.9. *Time and length are subject to Lorentz dilation and contraction*

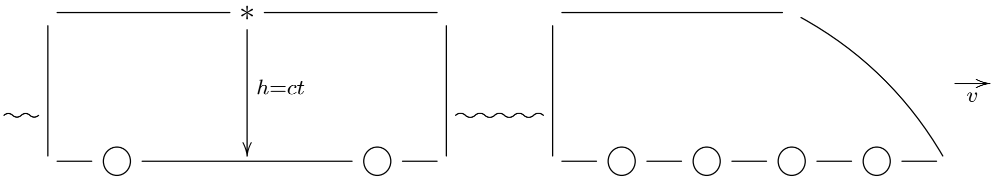
$$t\to\gamma t\quad,\quad L\to L/\gamma$$

where the number $\gamma\geq 1$, called Lorentz factor, is given by the formula

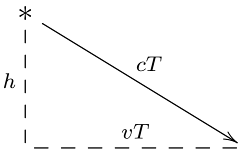
$$\gamma=\frac{1}{\sqrt{1-v^2/c^2}}$$

with v being the moving speed, at which time is measured.

PROOF. We use our favorite train. In order to compute the height h of the train, the passenger switches on the ceiling light bulb, measures the time t that the light needs to hit the floor, by traveling at speed c, and concludes that the train height is $h = ct$:



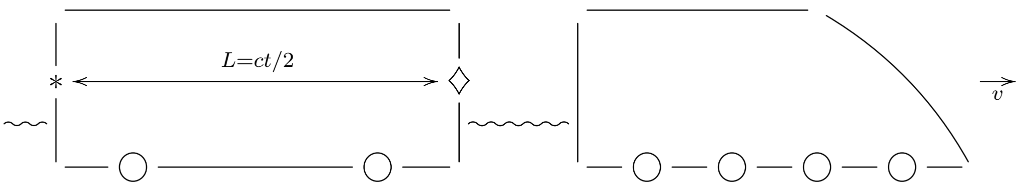
On the other hand, an observer on the ground will see here something different, namely a right triangle, with on the vertical the height of the train h, on the horizontal the distance vT that the train has traveled, and on the hypotenuse the distance cT that light has travelled, with T being the duration of the event, according to his watch:



But this gives, via Pythagoras and some calculus, the time dilation formula:

$$T = \gamma t$$

Regarding now length, imagine that the passenger wants to measure the length L of the car. For this purpose he switches on the light bulb, now at the rear of the car, and measures the time t needed for the light to reach the front of the car, and get reflected back by a mirror installed there, according to the following scheme:



Now viewed from the ground, the duration of the event is $T = T_1 + T_2$, where $T_1 > T_2$ are respectively the time needed for the light to travel forward, among others for beating v, and the time for the light to travel back, helped this time by v. More precisely, if l denotes the length of the train car viewed from the ground, the formula of T is:

$$T = \frac{l}{c-v} + \frac{l}{c+v} = \frac{2lc}{c^2 - v^2}$$

With this data, the formula $T = \gamma t$ of time dilation established before reads:

$$l = \frac{L}{\gamma}$$

Thus, we are led to the conclusions in the statement. $\square$

As a main consequence of the above, beautiful as they come, we have:

Theorem 1.10. *In the context of a relativistic object moving with speed v along the x axis, the frame change is given by the Lorentz transformation*

$$x' = \gamma(x - vt)$$

$$y' = y$$

$$z' = z$$

$$t' = \gamma(t - vx/c^2)$$

with $\gamma = 1/\sqrt{1 - v^2/c^2}$ *being as usual the Lorentz factor.*

Proof. We know that, with respect to the non-relativistic formulae, x is subject to the Lorentz dilation by γ, and so we obtain, as desired:

$$x' = \gamma(x - vt)$$

Regarding y, z, these are obviously unchanged, so done with these too. Finally, regarding time t, we can use here the reverse Lorentz transformation, given by:

$$x = \gamma(x' + vt')$$

$$y = y'$$

$$z = z'$$

By using the formula of x' we can compute t', and we obtain the following formula:

$$t' = \frac{x - \gamma x'}{\gamma v} = \frac{x - \gamma^2(x - vt)}{\gamma v} = \gamma\left(t - \frac{vx}{c^2}\right)$$

We are therefore led to the conclusion in the statement. □

1b. Maxwell equations

Getting now towards quantum mechanics, this was born in the early 20th century, from the study of the hydrogen atom, namely a negative charge, the electron, spinning around a positive charge, a proton. Which looks, at a first glance, like basic electrodynamics, and so it is with the Maxwell equations that we should start our study.

So, let us first talk about electrodynamics, and the Maxwell equations. We will see in a moment that these equations are indeed useful for getting started with hydrogen and quantum mechanics, but, and here comes the surprise, in a totally different way from the intuitive way one would expect. So, wait for it. Getting started now, we have:

Theorem 1.11. *Electrodynamics is governed by the formulae*

$$< \nabla, E >= \frac{\rho}{\varepsilon_0} \quad , \quad < \nabla, B >= 0$$

$$\nabla \times E = -\dot{B} \quad , \quad \nabla \times B = \mu_0 J + \mu_0 \varepsilon_0 \dot{E}$$

called Maxwell equations.

Proof. This is something fundamental, appearing as a tricky mixture of physics facts and mathematical results, the idea being as follows:

(1) To start with, electrodynamics is the science of moving electrical charges. And this is something quite complicated, because unlike in classical mechanics, where the Newton law is good for both the static and the dynamic setting, the Coulomb law, which is actually very similar to the Newton law, does the job when the charges are static, but no longer describes well the situation when the charges are moving.

(2) The problem comes from the fact that moving charges produce magnetism, and with this being visible when putting together two electric wires, which will attract or repel, depending on orientation. Thus, in contrast with classical mechanics, where static or dynamic problems are described by a unique field, the gravitational one, in electrodynamics we have two fields, namely the electric field E, and the magnetic field B.

(3) Fortunately, there is a full set of equations relating the electric field E and the magnetic field B, those above. Regarding the math, the dots denote derivatives with respect to time, and ∇ is the gradient operator, or space derivative, given by:

$$\nabla = \begin{pmatrix} \frac{d}{dx} \\ \frac{d}{dy} \\ \frac{d}{dz} \end{pmatrix}$$

(4) Regarding the physics, the first formula is the Gauss law, ρ being the charge, and ε_0 being a constant, and with this Gauss law more or less replacing the Coulomb law from electrostatics. The second formula is something basic, and anonymous. The third formula is the Faraday law. As for the fourth formula, this is the Ampère law, as modified by Maxwell, with J being the volume current density, and μ_0 being a constant. □

Quite surprisingly, the constants μ_0, ε_0 appearing above are related as follows:

Fact 1.12. *The constants* μ_0, ε_0 *are related by the Biot-Savart formula*

$$\mu_0 \varepsilon_0 = \frac{1}{c^2}$$

with $c = 299,792,458$ *being as usual the speed of light in vacuum.*

We will be back to this in a moment, with more about it. In the meantime, back to abstract electromagnetism, we have here the following key result, due to Lorentz:

Theorem 1.13. *The Maxwell equations found before, namely*

$$< \nabla, E >= \frac{\rho}{\varepsilon_0} \quad , \quad < \nabla, B >= 0$$

$$\nabla \times E = -\dot{B} \quad , \quad \nabla \times B = \mu_0 J + \mu_0 \varepsilon_0 \dot{E}$$

are invariant under Lorentz transformations.

Proof. Consider an electromagnetic field (E, B). This is altered by a Lorentz transformation into a field (E', B'), the equations for E' being as follows:

$$E'_x = E_x$$

$$E'_y = \gamma(E_y - vB_z)$$

$$E'_z = \gamma(E_z + vB_y)$$

As for the equations of B', these are quite similar, as follows:

$$B'_x = B_x$$

$$B'_y = \gamma\left(B_y + \frac{v}{c^2}E_z\right)$$

$$B'_z = \gamma\left(B_z - \frac{v}{c^2}E_y\right)$$

In order to do the math, consider the following matrices, with $\beta = v/c$ as usual:

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \end{pmatrix} \quad , \quad M = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\beta\gamma \\ 0 & \beta\gamma & 0 \end{pmatrix}$$

In terms of these matrices, the formulae for the new field (E', B') read:

$$E' = DE + cMB \quad , \quad B' = DB - \frac{M}{c}E$$

But this is already not that bad, and starting from these formulae, it is possible to prove that (E', B') satisfies as well the Maxwell equations, as desired. □

1c. Light, color, optics

Let us go back now to light. We first have the following basic result:

Theorem 1.14. *The wave equation in* $\mathbb{R}^N$ *is*

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

where Δ *is the Laplace operator, given by*

$$\Delta \varphi = \sum_{i=1}^N \frac{d^2 \varphi}{dx_i^2}$$

and $v > 0$ *is the propagation speed.*

Proof. As a first disclaimer, the equation in the statement is what comes out of experiments. However, allowing us a bit of imagination, and trust in this imagination, we can mathematically "prove" this equation, by discretizing, as follows:

(1) Let us first consider the 1D case. In order to understand the propagation of waves, we will model $\mathbb{R}$ as a network of balls, with springs between them, as follows:

$$\cdots\times\times\times\bullet\times\times\times\bullet\times\times\times\bullet\times\times\times\bullet\times\times\times\bullet\times\times\times\cdots$$

Now let us send an impulse, and see how the balls will be moving. For this purpose, we zoom on one ball. The situation here is as follows, l being the spring length:

$$\cdots\cdots\bullet_{\varphi(x-l)}\times\times\times\bullet_{\varphi(x)}\times\times\times\bullet_{\varphi(x+l)}\cdots\cdots$$

We have two forces acting at x. First is the Newton motion force, mass times acceleration, which is as follows, with m being the mass of each ball:

$$F_n = m\cdot\ddot{\varphi}(x)$$

And second is the Hooke force, displacement of the spring, times spring constant. Since we have two springs at x, this is as follows, k being the spring constant:

$$\begin{aligned}F_h &= F_h^r - F_h^l\\ &= k(\varphi(x+l)-\varphi(x)) - k(\varphi(x)-\varphi(x-l))\\ &= k(\varphi(x+l)-2\varphi(x)+\varphi(x-l))\end{aligned}$$

We conclude that the equation of motion, in our model, is as follows:

$$m\cdot\ddot{\varphi}(x) = k(\varphi(x+l)-2\varphi(x)+\varphi(x-l))$$

(2) Now let us take the limit of our model, as to reach to continuum. For this purpose we will assume that our system consists of $N>>0$ balls, having a total mass M, and spanning a total distance L. Thus, our previous infinitesimal parameters are as follows, with K being the spring constant of the total system, which is of course lower than k:

$$m=\frac{M}{N}\quad,\quad k=KN\quad,\quad l=\frac{L}{N}$$

With these changes, our equation of motion found in (1) reads:

$$\ddot{\varphi}(x) = \frac{KN^2}{M}(\varphi(x+l)-2\varphi(x)+\varphi(x-l))$$

Now observe that this equation can be written, more conveniently, as follows:

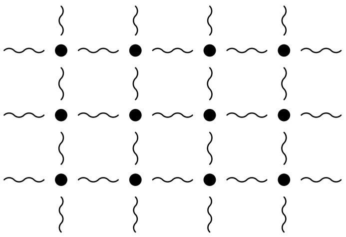
$$\ddot{\varphi}(x) = \frac{KL^2}{M}\cdot\frac{\varphi(x+l)-2\varphi(x)+\varphi(x-l)}{l^2}$$

With $N\to\infty$, and therefore $l\to 0$, we obtain in this way:

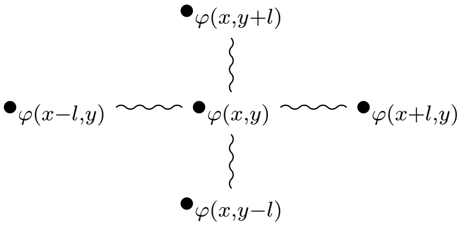
$$\ddot{\varphi}(x) = \frac{KL^2}{M}\cdot\frac{d^2\varphi}{dx^2}(x)$$

We are therefore led to the wave equation in the statement, which is $\ddot{\varphi} = v^2\varphi''$ in our present $N = 1$ dimensional case, the propagation speed being $v = \sqrt{K/M} \cdot L$.

(3) In 2 dimensions now, the same argument carries on. Indeed, we can use here a lattice model as follows, with all the edges standing for small springs:



As before in one dimension, we send an impulse, and we zoom on one ball. The situation here is as follows, with l being the spring length:



By doing the math, as before in 1D, we are led to the following equation:

$$\begin{aligned}\ddot{\varphi}(x,y) &= \frac{KL^2}{M} \times \frac{\varphi(x+l,y) - 2\varphi(x,y) + \varphi(x-l,y)}{l^2}\\ &+ \frac{KL^2}{M} \times \frac{\varphi(x,y+l) - 2\varphi(x,y) + \varphi(x,y-l)}{l^2}\end{aligned}$$

With $N \to \infty$, and therefore $l \to 0$, we obtain in this way:

$$\ddot{\varphi}(x,y) = \frac{KL^2}{M}\left(\frac{d^2\varphi}{dx^2} + \frac{d^2\varphi}{dy^2}\right)(x,y)$$

Thus, we are led in this way to the wave equation in two dimensions, as in the statement, with $v = \sqrt{K/M} \cdot L$ being the propagation speed of our wave.

(4) In 3 dimensions now, which is the case of the main interest, corresponding to our real-life world, the same argument carries over, and the wave equation is as follows:

$$\ddot{\varphi}(x,y,z) = v^2\left(\frac{d^2\varphi}{dx^2} + \frac{d^2\varphi}{dy^2} + \frac{d^2\varphi}{dz^2}\right)(x,y,z)$$

Finally, the same argument carries on in arbitrary N dimensions. $\square$

The point now is that, in relation with the Maxwell equations, we have:

Theorem 1.15. *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.

Proof. Under the circumstances in the statement, namely no charge or current present, the Maxwell equations in Theorem 1.11 simply read:

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

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

Now by applying the curl operator to the last two equations, we obtain:

$$\nabla \times (\nabla \times E) = -\nabla \times \dot{B} = -(\nabla \times B)' = -\ddot{E}/c^2$$

$$\nabla \times (\nabla \times B) = \nabla \times \dot{E}/c^2 = (\nabla \times E)'/c^2 = -\ddot{B}/c^2$$

But the double curl operator is subject to the following formula:

$$\nabla \times (\nabla \times \varphi) = \nabla < \nabla, \varphi > - \Delta \varphi$$

Now by using the first two equations, we are led to the conclusion in the statement. □

So, what is light? Light is the wave predicted by Theorem 1.15, traveling in vacuum at the maximum possible speed, c, 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 1.16. *An accelerating or decelerating charge produces electromagnetic radiation, called light, whose frequency and wavelength can be explicitly 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.

Moving ahead, let us go back to the wave equation $\ddot{\varphi} = v^2 \Delta \varphi$ from Theorem 1.14, and try to understand its simplest solutions. In 1D, the situation is as follows:

Theorem 1.17. *The 1D wave equation has as basic solutions the 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. There are several things going on here, the idea being as follows:

(1) 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 $v=w/k$:

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

(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. As for the last assertion, this 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 1.18. *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 1.17, 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 1.17. And second is the assumption, in connection with the Fourier decomposition result from the end of Theorem 1.17, 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 Theorem 1.15 and Fact 1.16, we have:

Fact 1.19. *Physically speaking, a monochromatic plane wave is the electromagnetic radiation appearing as in Theorem 1.15 and Fact 1.16, 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 1.17 and Definition 1.18, into the context of our original physics discussion, from Fact 1.16. And also to add an extra twist coming from physics, and more specifically from the notion of polarization. More on this later.

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	$625 - 750$

Back now to our business, with the above theory of light 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 1.20. *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 statement is precisely the first part of Theorem 1.15, with the vacuum constants ε_0, μ_0 being now replaced by their versions ε, μ, concerning the medium in question. In what regards now the second part of Theorem 1.15, we have:

Theorem 1.21. *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, from the above, and the proof in general is identical, with the resulting speed being:

$$v = \frac{1}{\sqrt{\varepsilon\mu}}$$

But this formula can be written is a more familiar from, as above. □

Next in line, and of interest for us, as we will soon discover, we have:

Fact 1.22. *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, of great practical interest, we have:

Theorem 1.23. *Light can be decomposed, by using a prism.*

Proof. This follows from Fact 1.22. 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. □

With this in hand, we can now talk about spectroscopy:

Fact 1.24. *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 it, 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.

1d. Atoms, Bohr model

Getting now to some truly exciting applications of light and spectroscopy, let us discuss the beginnings of the atomic theory. There is a long story here, involving many discoveries, around 1890-1900, focusing on hydrogen H. We will present here things a bit retrospectively. First on our list is the following discovery, by Lyman in 1906:

Fact 1.25 (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, which is invisible to the naked eye. Due to this fact, this series, while theoretically being the most important, 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 1.26 (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. As a third main result now, this time in IR, due to Paschen in 1908, we have:

FACT 1.27 (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
⋮	⋮	⋮	⋮
∞	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 1.28 (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.

From a theoretical physics viewpoint, the main result remains the middle assertion, called Ritz-Rydberg combination principle, which is something quite puzzling. But this combination principle reminds the formula $e_{n_1n_2}e_{n_2n_3} = e_{n_1n_3}$ for the usual matrix units $e_{ij} : e_j \to e_i$. Thus, we are in familiar territory here, and we can start dreaming of:

THOUGHT 1.29. *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.*

Now back to more concrete things, 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 for all this, let us agree on:

CLAIM 1.30. *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, as a natural question, 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 it is so came from several clever experiments. Let us first recall that careful experiments with tiny particles are practically impossible. However, all sorts of brutal experiments, such as bombarding matter with other pieces of matter, accelerated to the extremes, or submitting it to huge electric and magnetic fields, do work. And it is such kind of experiments, due to Thomson, Rutherford and others, "peeling off" protons $+$, neutrons 0 and electrons $-$ from matter, and observing them, that led to the conclusion that these small beasts $+,0,-$ exist indeed, in agreement with Claim 1.30.

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 ${}_{92}U$, polonium ${}_{84}Po$ and radium ${}_{88}Ra$. 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 1.30, 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 1.30.

So, taking now Claim 1.30 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 "brutal" experiments of Thomson, Rutherford and others, which not only proved Claim 1.30, but led to an improved version of it, as follows:

Claim 1.31. *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. And with this in mind, all the pieces of our puzzle start fitting together, and we are led to the following grand conclusion:

Claim 1.32 (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 mathematics in all the above can lead to a beautiful axiomatization of quantum mechanics, as a "matrix mechanics", along the lines of Thought 1.29.

1e. Exercises

Exercises:

EXERCISE 1.33.

EXERCISE 1.34.

EXERCISE 1.35.

EXERCISE 1.36.

EXERCISE 1.37.

EXERCISE 1.38.

EXERCISE 1.39.

EXERCISE 1.40.

Bonus exercise.

CHAPTER 2

Schrödinger equation

2a. Schrödinger equation

We have seen in the previous chapter that the spectral data of the hydrogen atom, as recorded by Lyman, Balmer and Paschen, led to the Ritz-Rydberg combination principle, whose mathematics reminds the matrix multiplication, and which leads to the conclusion that quantum mechanics should be some sort of “matrix mechanics”.

We will see in a moment, following Heisenberg, that this is indeed true. However, before getting into this, let us hear as well the point of view of Schrödinger, which came a few years later. The idea of Schrödinger was to 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:

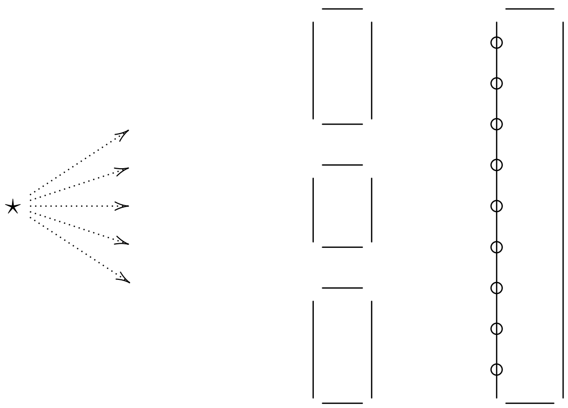
$$\varphi_t(x)=\varphi_0(x)+\gamma(t)$$

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*,*

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

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.

2b. Basic properties

With the above discussed, and more on this, general theory, on several occasions, in what follows, and with the promise here of course to come back soon to Heisenberg too, 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 Leibniz product rule for the derivatives, we have the following formula, for the time derivative of the probability density function:

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

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. □

2c. Position, momentum

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.

Finally, still at the level of generalities, if there is one classical equation which reminds the Schrödinger one, that is the heat equation, which appears as follows:

Theorem 2.18. *Heat diffusion in $\mathbb{R}^N$ is described by the heat equation*

$$\dot{\varphi} = \alpha\Delta\varphi$$

where $\alpha > 0$ is the thermal diffusivity of the medium, and Δ is the Laplace operator.

Proof. The study here is quite similar to the study of waves, as follows:

(1) To start with, as an intuitive explanation for the equation, since the second derivative φ'' in one dimension, or the quantity $\Delta\varphi$ in general, computes the average value of a function φ around a point, minus the value of φ at that point, the heat equation as formulated above tells us that the rate of change $\dot{\varphi}$ of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha > 0$, to the average difference of temperature between that given point and the surrounding material.

(2) The heat equation as formulated above is of course something approximative, and several improvements can be made to it, first by incorporating a term accounting for heat radiation, and then doing several fine-tunings, depending on the material involved. But more on this later, for the moment let us focus on the heat equation above.

(3) The idea is that we can recover this equation a bit as we did for the wave equation in chapter 1, by using a basic lattice model. Indeed, let us first assume that we are in the case $N = 1$. Here our model looks as follows, with distance $l > 0$ between neighbors:

$$-\!\!-\!\!-\ \circ_{x-l} \overset{l}{-\!\!-\!\!-} \circ_x \overset{l}{-\!\!-\!\!-} \circ_{x+l}\ -\!\!-\!\!-$$

In order to model heat diffusion, we have to implement the intuitive mechanism explained above, namely “the rate of change of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha > 0$, to the average difference of temperature between that given point and the surrounding material”.

(4) In practice, this leads to a condition as follows, expressing the change of the temperature φ, over a small period of time $\delta > 0$:

$$\varphi(x,t+\delta)=\varphi(x,t)+\frac{\alpha\delta}{l^2}\sum_{x\sim y}\left[\varphi(y,t)-\varphi(x,t)\right]$$

To be more precise, we have made several assumptions here, as follows:

– General heat diffusion assumption: the change of temperature at any given point x is proportional to the average over neighbors, $y \sim x$, of the differences $\varphi(y,t)-\varphi(x,t)$ between the temperatures at x, and at these neighbors y.

– Infinitesimal time and length conditions: in our model, the change of temperature at a given point x is proportional to small period of time involved, $\delta > 0$, and is inverse proportional to the square of the distance between neighbors, l^2.

(5) Regarding these latter assumptions, the one regarding the proportionality with the time elapsed $\delta > 0$ is something quite natural, physically speaking, and mathematically speaking too, because we can rewrite our equation as follows, making it clear that we have here an equation regarding the rate of change of temperature at x:

$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\frac{\alpha}{l^2}\sum_{x\sim y}\left[\varphi(y,t)-\varphi(x,t)\right]$$

As for the second assumption that we made above, namely inverse proportionality with l^2, this can be justified on physical grounds too, but again, perhaps the best is to do the math, which will show right away where this proportionality comes from.

(6) So, let us do the math. In the context of our 1D model the neighbors of x are the points $x \pm l$, and so the equation that we wrote above takes the following form:

$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\frac{\alpha}{l^2}\Big[(\varphi(x+l,t)-\varphi(x,t))+(\varphi(x-l,t)-\varphi(x,t))\Big]$$

Now observe that we can write this equation as follows:

$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\alpha\cdot\frac{\varphi(x+l,t)-2\varphi(x,t)+\varphi(x-l,t)}{l^2}$$

(7) As it was the case with the wave equation in chapter 1, we recognize on the right the usual approximation of the second derivative, coming from calculus. Thus, when taking the continuous limit of our model, $l \to 0$, we obtain the following equation:

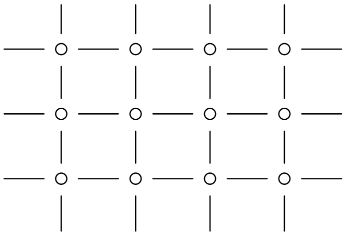
$$\frac{\varphi(x,t+\delta)-\varphi(x,t)}{\delta}=\alpha\cdot\varphi''(x,t)$$

Now with $t \to 0$, we are led in this way to the 1D heat equation, namely:

$$\dot{\varphi}(x,t) = \alpha \cdot \varphi''(x,t)$$

(8) In practice now, there are of course still a few details to be discussed, in relation with all this, for instance at the end, in relation with the precise order of the limiting operations $l \to 0$ and $\delta \to 0$ to be performed, but these remain minor aspects, because our equation makes it clear, right from the beginning, that time and space are separated, and so that there is no serious issue with all this. And so, fully done with 1D.

(9) With this done, let us discuss now 2 dimensions. Here, as before for the waves, we can use a lattice model as follows, with all lengths being $l > 0$, for simplifying:



(10) We have to implement now the physical heat diffusion mechanism, namely "the rate of change of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha > 0$, to the average difference of temperature between that given point and the surrounding material". In practice, this leads to a condition as follows, expressing the change of the temperature φ, over a small period of time $\delta > 0$:

$$\varphi(x,y,t+\delta) = \varphi(x,y,t) + \frac{\alpha\delta}{l^2} \sum_{(x,y)\sim(u,v)} [\varphi(u,v,t) - \varphi(x,y,t)]$$

In fact, we can rewrite our equation as follows, making it clear that we have here an equation regarding the rate of change of temperature at x:

$$\frac{\varphi(x,y,t+\delta) - \varphi(x,y,t)}{\delta} = \frac{\alpha}{l^2} \sum_{(x,y)\sim(u,v)} [\varphi(u,v,t) - \varphi(x,y,t)]$$

(11) So, let us do the math. In the context of our 2D model the neighbors of x are the points $(x \pm l, y \pm l)$, so the equation above takes the following form:

$$\begin{aligned} & \frac{\varphi(x,y,t+\delta) - \varphi(x,y,t)}{\delta} \\ = \ & \frac{\alpha}{l^2}\Big[(\varphi(x+l,y,t) - \varphi(x,y,t)) + (\varphi(x-l,y,t) - \varphi(x,y,t))\Big] \\ + \ & \frac{\alpha}{l^2}\Big[(\varphi(x,y+l,t) - \varphi(x,y,t)) + (\varphi(x,y-l,t) - \varphi(x,y,t))\Big] \end{aligned}$$

Now observe that we can write this equation as follows:

$$\begin{aligned}\frac{\varphi(x,y,t+\delta)-\varphi(x,y,t)}{\delta} &= \alpha\cdot\frac{\varphi(x+l,y,t)-2\varphi(x,y,t)+\varphi(x-l,y,t)}{l^2}\\ &+ \alpha\cdot\frac{\varphi(x,y+l,t)-2\varphi(x,y,t)+\varphi(x,y-l,t)}{l^2}\end{aligned}$$

(12) As it was the case when modelling the wave equation in chapter 1, we recognize on the right the usual approximation of the second derivative, coming from calculus. Thus, when taking the continuous limit of our model, $l\to 0$, we obtain the following equation:

$$\frac{\varphi(x,y,t+\delta)-\varphi(x,y,t)}{\delta}=\alpha\left(\frac{d^2\varphi}{dx^2}+\frac{d^2\varphi}{dy^2}\right)(x,y,t)$$

Now with $t\to 0$, we are led in this way to the heat equation, namely:

$$\dot{\varphi}(x,y,t)=\alpha\cdot\Delta\varphi(x,y,t)$$

Finally, in arbitrary N dimensions the same argument carries over, namely a straightforward lattice model, and gives the heat equation, as formulated in the statement. □

2d. Separation of variables

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 2.19. *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 2.20. *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.

Moving ahead with theory, we can further build on Theorem 2.19, with a number of key observations on the time-independent Schrödinger equation, as follows:

Theorem 2.21. *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 2.22. *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 2.23. *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”.

2e. Exercises

Exercises:

Exercise 2.24.

Exercise 2.25.

Exercise 2.26.

Exercise 2.27.

Exercise 2.28.

Exercise 2.29.

Exercise 2.30.

Exercise 2.31.

Bonus exercise.

CHAPTER 3

Quantum mechanics

3a. Hilbert spaces

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:

$$4Re(w<x,y>)^2-4||x||^2\cdot||y||^2\leq0$$

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*

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

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.

3b. Linear operators

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). $\square$

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:

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

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 third and last basic 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. □

Let us discuss now the diagonalization problem for the operators $T \in B(H)$, in analogy with the diagonalization problem for the usual matrices $A \in M_N(\mathbb{C})$. We first have:

Definition 3.10. *The spectrum of an operator $T \in B(H)$ is the set*

$$\sigma(T) = \left\{ \lambda \in \mathbb{C} \Big| T - \lambda \notin B(H)^{-1} \right\}$$

where $B(H)^{-1} \subset B(H)$ is the set of invertible operators.

As a basic example, in the finite dimensional case, $H=\mathbb{C}^N$, the spectrum of a usual matrix $A\in M_N(\mathbb{C})$ is the collection of its eigenvalues, taken without multiplicities. We will see many other examples. In general, the spectrum has the following properties:

Proposition 3.11. *The spectrum of $T\in B(H)$ contains the eigenvalue set*

$$\varepsilon(T)=\left\{\lambda\in\mathbb{C}\Big|\ker(T-\lambda)\neq\{0\}\right\}$$

and $\varepsilon(T)\subset\sigma(T)$ is an equality in finite dimensions, but not in infinite dimensions.

Proof. We have several assertions here, the idea being as follows:

(1) First of all, the eigenvalue set is indeed the one in the statement, because $Tx=\lambda x$ tells us precisely that $T-\lambda$ must be not injective. The fact that we have $\varepsilon(T)\subset\sigma(T)$ is clear as well, because if $T-\lambda$ is not injective, it is not bijective.

(2) In finite dimensions we have $\varepsilon(T)=\sigma(T)$, because $T-\lambda$ is injective if and only if it is bijective, with the boundedness of the inverse being automatic.

(3) In infinite dimensions we can assume $H=l^2(\mathbb{N})$, and the shift operator $S(e_i)=e_{i+1}$ is injective but not surjective. Thus $0\in\sigma(T)-\varepsilon(T)$. ☐

Philosophically, the best way of thinking at this is as follows: the numbers $\lambda\notin\sigma(T)$ are good, because we can invert $T-\lambda$, the numbers $\lambda\in\sigma(T)-\varepsilon(T)$ are bad, because so they are, and the eigenvalues $\lambda\in\varepsilon(T)$ are evil. Welcome to operator theory.

Let us develop now some general theory. Here is a first basic result regarding the spectra, inspired from what happens in finite dimensions, and which shows that things do not necessarily extend without troubles to the infinite dimensional setting:

Theorem 3.12. *We have the following formula, valid for any operators S,T:*

$$\sigma(ST)\cup\{0\}=\sigma(TS)\cup\{0\}$$

In finite dimensions we have $\sigma(ST)=\sigma(TS)$, but this fails in infinite dimensions.

Proof. There are several assertions here, the idea being as follows:

(1) Let us first prove the main assertion, stating that $\sigma(ST),\sigma(TS)$ coincide outside 0. We first prove that we have the following implication:

$$1\notin\sigma(ST)\implies 1\notin\sigma(TS)$$

Assume indeed that $1-ST$ is invertible, with inverse denoted R:

$$R=(1-ST)^{-1}$$

We have then the following formulae, relating our variables R,S,T:

$$RST=STR=R-1$$

By using $RST = R - 1$, we have the following computation:

$$\begin{aligned}(1+TRS)(1-TS) &= 1+TRS-TS-TRSTS\\ &= 1+TRS-TS-TRS+TS\\ &= 1\end{aligned}$$

A similar computation, using $STR = R - 1$, shows that we have:

$$(1-TS)(1+TRS) = 1$$

Thus $1 - TS$ is invertible, with inverse $1 + TRS$, which proves our claim. Now by multiplying by scalars, we deduce that for any $\lambda \in \mathbb{C} - \{0\}$ we have, as desired:

$$\lambda \notin \sigma(ST) \implies \lambda \notin \sigma(TS)$$

(2) Regarding now the counterexample to the formula $\sigma(ST) = \sigma(TS)$, in general, let us take S to be the shift on $H = L^2(\mathbb{N})$, given by the following formula:

$$S(e_i) = e_{i+1}$$

As for T, we can take it to be the adjoint of S, which is the following operator:

$$S^*(e_i) = \begin{cases} e_{i-1} & \text{if } i > 0\\ 0 & \text{if } i = 0\end{cases}$$

Let us compose now these two operators. In one sense, we have:

$$S^*S = 1 \implies 0 \notin \sigma(SS^*)$$

In the other sense, however, the situation is different, as follows:

$$SS^* = Proj(e_0^\perp) \implies 0 \in \sigma(SS^*)$$

Thus, the spectra do not match on 0, and we have our counterexample, as desired. □

Let us develop now some systematic theory for the computation of the spectra, based on what we know about the eigenvalues of the usual complex matrices. As a first result, which is well-known for the usual matrices, and extends well, we have:

Theorem 3.13. *We have the "rational functional calculus" formula*

$$\sigma(f(T)) = f(\sigma(T))$$

valid for any rational function $f \in \mathbb{C}(X)$ having poles outside $\sigma(T)$.

Proof. This can be proved in two steps, as follows:

(1) Assume first that our rational function $f \in \mathbb{C}(X)$ is a usual polynomial $P \in \mathbb{C}[X]$. We pick a scalar $\lambda \in \mathbb{C}$, and we decompose the polynomial $P - \lambda$, as follows:

$$P(X) - \lambda = c(X - r_1)\ldots(X - r_n)$$

We have then the following equivalences, which give the result:

$$\begin{aligned}\lambda\notin\sigma(P(T)) &\iff P(T)-\lambda\in B(H)^{-1}\\ &\iff c(T-r_1)\ldots(T-r_n)\in B(H)^{-1}\\ &\iff T-r_1,\ldots,T-r_n\in B(H)^{-1}\\ &\iff r_1,\ldots,r_n\notin\sigma(T)\\ &\iff \lambda\notin P(\sigma(T))\end{aligned}$$

(2) In general now, we pick a scalar $\lambda\in\mathbb{C}$, we write $f=P/Q$, and we set $F=P-\lambda Q$. By using what we found in (1), for this polynomial $F\in\mathbb{C}[X]$, we obtain:

$$\begin{aligned}\lambda\in\sigma(f(T)) &\iff F(T)\notin B(H)^{-1}\\ &\iff 0\in\sigma(F(T))\\ &\iff 0\in F(\sigma(T))\\ &\iff \exists\mu\in\sigma(T),F(\mu)=0\\ &\iff \lambda\in f(\sigma(T))\end{aligned}$$

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

As a first application of the above methods, we have the following key result:

Theorem 3.14. *The following happen:*

1. *For a unitary operator, $U^*=U^{-1}$, we have $\sigma(U)\subset\mathbb{T}$.*
2. *For a self-adjoint operator, $T=T^*$, we have $\sigma(T)\subset\mathbb{R}$.*

Proof. This is something quite tricky, based on Theorem 3.13, as follows:

(1) Assuming $U^*=U^{-1}$, we have the following norm computation:

$$||U||=\sqrt{||UU^*||}=\sqrt{1}=1$$

Now if we denote by D the unit disk, we obtain from this:

$$\sigma(U)\subset D$$

On the other hand, once again by using $U^*=U^{-1}$, we have as well:

$$||U^{-1}||=||U^*||=||U||=1$$

Thus, as before with D being the unit disk in the complex plane, we have:

$$\sigma(U^{-1})\subset D$$

Now by using Theorem 3.13, we obtain $\sigma(U)\subset D\cap D^{-1}=\mathbb{T}$, as desired.

(2) Consider the following rational function, depending on a parameter $r\in\mathbb{R}$:

$$f(z)=\frac{z+ir}{z-ir}$$

Then for $r >> 0$ the operator $f(T)$ is well-defined, and we have:

$$\left(\frac{T+ir}{T-ir}\right)^* = \frac{T-ir}{T+ir} = \left(\frac{T+ir}{T-ir}\right)^{-1}$$

Thus $f(T)$ is unitary, and by (1) we have $\sigma(T) \subset f^{-1}(\mathbb{T}) = \mathbb{R}$, as desired. □

In order to formulate our next result, we will need the following notion:

Definition 3.15. *Given an operator $T \in B(H)$, its spectral radius*

$$\rho(T) \in \big[0, ||T||\big]$$

is the radius of the smallest disk centered at 0 containing $\sigma(T)$.

Now with this notion in hand, we have the following key result:

Theorem 3.16. *The spectral radius of an operator $T \in B(H)$ is given by*

$$\rho(T) = \lim_{n\to\infty} ||T^n||^{1/n}$$

and in this formula, we can replace the limit by an inf.

Proof. We have several things to be proved, the idea being as follows:

(1) Our first claim is that the numbers $u_n = ||T^n||^{1/n}$ satisfy:

$$(n+m)u_{n+m} \leq nu_n + mu_m$$

Indeed, we have the following estimate, using the Young inequality $ab \leq a^p/p + b^q/q$, with exponents $p = (n+m)/n$ and $q = (n+m)/m$:

$$\begin{aligned}
u_{n+m} &= ||T^{n+m}||^{1/(n+m)}\\
&\leq ||T^n||^{1/(n+m)}||T^m||^{1/(n+m)}\\
&\leq ||T^n||^{1/n} \cdot \frac{n}{n+m} + ||T^m||^{1/m} \cdot \frac{m}{n+m}\\
&= \frac{nu_n + mu_m}{n+m}
\end{aligned}$$

(2) Our second claim is that the second assertion holds, namely:

$$\lim_{n\to\infty} ||T^n||^{1/n} = \inf_n ||T^n||^{1/n}$$

For this purpose, we just need the inequality found in (1). Indeed, fix $m \geq 1$, let $n \geq 1$, and write $n = lm + r$ with $0 \leq r \leq m-1$. By using twice $u_{ab} \leq u_b$, we get:

$$u_n \leq \frac{1}{n}(lmu_{lm} + ru_r) \leq u_m + \frac{r}{n}u_1$$

It follows that we have $\limsup_n u_n \leq u_m$, which proves our claim.

(3) Summarizing, we are left with proving the main formula, which is as follows, and with the remark that we already know that the sequence on the right converges:

$$\rho(T)=\lim_{n\to\infty}||T^n||^{1/n}$$

In one sense, we can use the polynomial calculus formula $\sigma(T^n)=\sigma(T)^n$. Indeed, this gives the following estimate, valid for any n, as desired:

$$\begin{aligned}\rho(T)&=\sup_{\lambda\in\sigma(T)}|\lambda|\\&=\sup_{\rho\in\sigma(T)^n}|\rho|^{1/n}\\&=\sup_{\rho\in\sigma(T^n)}|\rho|^{1/n}\\&=\rho(T^n)^{1/n}\\&\leq||T^n||^{1/n}\end{aligned}$$

(4) For the reverse inequality, we fix a number $\rho>\rho(T)$, and we want to prove that we have $\rho\geq\lim_{n\to\infty}||T^n||^{1/n}$. By using the Cauchy formula, we have:

$$\begin{aligned}\frac{1}{2\pi i}\int_{|z|=\rho}\frac{z^n}{z-T}\,dz&=\frac{1}{2\pi i}\int_{|z|=\rho}\sum_{k=0}^\infty z^{n-k-1}T^k\,dz\\&=\sum_{k=0}^\infty\frac{1}{2\pi i}\left(\int_{|z|=\rho}z^{n-k-1}dz\right)T^k\\&=\sum_{k=0}^\infty\delta_{n,k+1}T^k\\&=T^{n-1}\end{aligned}$$

By applying the norm we obtain from this formula:

$$\begin{aligned}||T^{n-1}||&\leq\frac{1}{2\pi}\int_{|z|=\rho}\left|\left|\frac{z^n}{z-T}\right|\right|\,dz\\&\leq\rho^n\cdot\sup_{|z|=\rho}\left|\left|\frac{1}{z-T}\right|\right|\end{aligned}$$

Since the sup does not depend on n, by taking n-th roots, we obtain in the limit:

$$\rho\geq\lim_{n\to\infty}||T^n||^{1/n}$$

Now recall that ρ was by definition an arbitrary number satisfying $\rho>\rho(T)$. Thus, we have proved the reverse inequality, and we are led to the conclusion in the statement. □

In the case of the normal elements, we have the following finer result:

Theorem 3.17. *The spectral radius of a normal element,*

$$TT^*=T^*T$$

is equal to its norm.

Proof. We can proceed in two steps, as follows:

Step 1. In the case $T=T^*$ we have $||T^n||=||T||^n$ for any exponent of the form $n=2^k$, by using the formula $||TT^*||=||T||^2$, and by taking n-th roots we get:

$$\rho(T)\geq||T||$$

Thus, we are done with the self-adjoint case, with the result $\rho(T)=||T||$.

Step 2. In the general normal case $TT^*=T^*T$ we have $T^n(T^n)^*=(TT^*)^n$, and by using this, along with the result from Step 1, applied to TT^*, we obtain:

$$\begin{aligned}\rho(T)&=\lim_{n\to\infty}||T^n||^{1/n}\\&=\sqrt{\lim_{n\to\infty}||T^n(T^n)^*||^{1/n}}\\&=\sqrt{\lim_{n\to\infty}||(TT^*)^n||^{1/n}}\\&=\sqrt{\rho(TT^*)}\\&=\sqrt{||T||^2}\\&=||T||\end{aligned}$$

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

It is possible to further build on the above, in a quite straightforward, but long and technical way, as to have in the end a spectral theorem for the normal operators:

Theorem 3.18. *The following happen:*

1. *Any self-adjoint operator, $T=T^*$, is diagonalizable.*
2. *More generally, any normal operator, $TT^*=T^*T$, is diagonalizable.*
3. *In fact, any family $\{T_i\}$ of commuting normal operators is diagonalizable.*

Proof. This is certainly a tough theorem, with (1,2,3) coming by generalizing the Spectral Theorem, in its various incarnations, for the usual matrices $M\in M_N(\mathbb{C})$. □

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].

3c. Quantum mechanics

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

Axioms 3.19. *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.

3d. Uncertainty principle

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

Theorem 3.20 (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].

3e. Exercises

Exercises:

EXERCISE 3.21.

EXERCISE 3.22.

EXERCISE 3.23.

EXERCISE 3.24.

EXERCISE 3.25.

EXERCISE 3.26.

EXERCISE 3.27.

EXERCISE 3.28.

Bonus exercise.

CHAPTER 4

Hydrogen atom

4a. Spherical coordinates

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 4.1. *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:

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

(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:

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

(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:

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

(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 $(A^{-1})^tA^{-1}$ 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:

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

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 4.2. *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 4.1, 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. □

4b. Angular equation

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 4.3. *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. $\square$

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

Proposition 4.4. *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. $\square$

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 4.5. *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:

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

But this is the Legendre equation, as stated. □

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

Theorem 4.6. *The solutions of the Legendre equation, namely*

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

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 4.6, 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 4.7. *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 ± 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 4.8. *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 4.9. *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. □

4c. Hydrogen atom

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 4.10. *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 4.11. *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.

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

Fact 4.12. *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 4.13 (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 4.14. *The Rydberg constant for hydrogen is given by*

$$R=-\frac{E_1}{h_0c}$$

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

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

simply reads, via the energy formula in Theorem 4.13,

$$\frac{1}{\lambda_{n_1n_2}}=\frac{E_{n_2}-E_{n_1}}{h_0c}$$

which is in agreement with the Planck formula $E=h_0c/\lambda$.

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

$$R=\frac{-E_1}{h_0c}=\frac{2.177\times 10^{-18}}{6.626\times 10^{-34}\times 2.998\times 10^8}=1.096\times 10^7$$

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

$$R=\frac{m}{2h_0c}\left(\frac{Ke^2}{h}\right)^2$$

Regarding now the second assertion, by dividing $R = -E_1/(h_0c)$ by any number of type n^2 we obtain, according to the energy convention in Theorem 4.13:

$$\frac{R}{n^2} = -\frac{E_n}{h_0c}$$

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. $\square$

4d. Wave functions

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

Theorem 4.15. *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. $\square$

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

Theorem 4.16. *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 4.13, 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 4.15 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 4.15, the point is that the polynomials $v(p)$ appearing there are well-known objects in mathematics, as follows:

Proposition 4.17. *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. $\square$

With the above result in hand, we can now improve Theorem 4.15, as follows:

THEOREM 4.18. *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 4.15 and Proposition 4.17, and then doing some remaining work, concerning the normalization of the wave function, which leads to the normalization factor appearing above. $\square$

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.

4e. Exercises

Exercises:

EXERCISE 4.19.

EXERCISE 4.20.

EXERCISE 4.21.

EXERCISE 4.22.

EXERCISE 4.23.

EXERCISE 4.24.

EXERCISE 4.25.

EXERCISE 4.26.

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. Discussion

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.

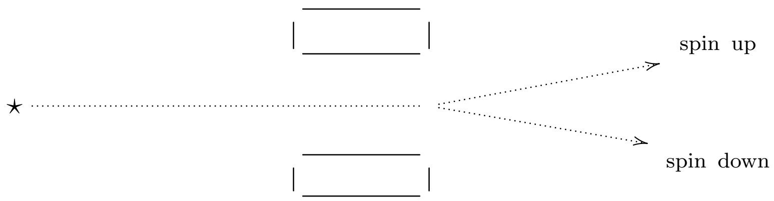
So, as a first question that we would like to solve, we have:

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. 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.

5b. Rotating objects

Getting started now, let us first talk about rotating objects in the context of classical mechanics. We will need the following notion, that we well know from before:

Definition 5.8. *The vector product of two vectors in* $\mathbb{R}^3$ *is given by*

$$x\times y=||x||\cdot||y||\cdot\sin\theta\cdot n$$

where $n\in\mathbb{R}^3$ *with* $n\perp x,y$ *and* $||n||=1$ *is constructed using the right-hand rule:*

$$\begin{matrix}\quad\uparrow_{x\times y}\\ \leftarrow_x\quad\\ \swarrow_y\quad\end{matrix}$$

Alternatively, in usual vertical linear algebra notation for all vectors,

$$\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}\times\begin{pmatrix}y_1\\ y_2\\ y_3\end{pmatrix}=\begin{pmatrix}x_2y_3-x_3y_2\\ x_3y_1-x_1y_3\\ x_1y_2-x_2y_1\end{pmatrix}$$

the rule being that of computing 2×2 *determinants, and adding a middle sign.*

Obviously, this definition is something quite subtle, and also something very annoying, because you always need this, and always forget the formula. Here are my personal methods. With the first definition, what I always remember is that:

$$||x\times y||\sim ||x||,||y||\quad,\quad x\times x=0\quad,\quad e_1\times e_2=e_3$$

So, here's how it works. We are looking for a vector $x\times y$ whose length is proportional to those of x,y. But the second formula tells us that the angle θ between x,y must be involved via $0\to 0$, and so the factor can only be $\sin\theta$. And with this we are almost there, it's just a matter of choosing the orientation, and this comes from $e_1\times e_2=e_3$.

As with the second definition, that I like the most, what I remember here is simply:

$$\begin{vmatrix}1&x_1&y_1\\1&x_2&y_2\\1&x_3&y_3\end{vmatrix}=?$$

Indeed, when trying to compute this determinant, by developing over the first column, what you get as coefficients are the entries of $x\times y$. And with the good middle sign.

In practice now, in order to get familiar with the vector products, nothing better than doing some classical mechanics. We have here the following key result:

Theorem 5.9. *In the gravitational 2-body problem, the angular momentum*

$$J=x\times p$$

with $p=mv$ being the usual momentum, is conserved.

Proof. There are several things to be said here, the idea being as follows:

(1) First of all the usual momentum, $p=mv$, is not conserved, because the simplest solution is the circular motion, where the moment gets turned around. But this suggests precisely that, in order to fix the lack of conservation of the momentum p, what we have to do is to make a vector product with the position x. Leading to J, as above.

(2) Regarding now the proof, consider indeed a particle m moving under the gravitational force of a particle M, assumed, as usual, to be fixed at 0. By using the fact that for two proportional vectors, $p\sim q$, we have $p\times q=0$, we obtain:

$$\begin{aligned}\dot{J}&=\dot{x}\times p+x\times\dot{p}\\&=v\times mv+x\times ma\\&=m(v\times v+x\times a)\\&=m(0+0)\\&=0\end{aligned}$$

Now since the derivative of J vanishes, this quantity is constant, as stated. □

While the above principle looks like something quite trivial, the mathematics behind it is quite interesting, and has several notable consequences, as follows:

THEOREM 5.10. *In the context of a 2-body problem, the following happen:*

1. *The fact that the direction of J is fixed tells us that the trajectory of one body with respect to the other lies in a plane.*
2. *The fact that the magnitude of J is fixed tells us that the Kepler 2 law holds, namely that we have same areas sweeped by Ox over the same times.*

PROOF. This follows indeed from Theorem 5.9, as follows:

(1) We have by definition $J = m(x \times v)$, and since a vector product is orthogonal on both the vectors it comes from, we deduce from this that we have:

$$J \perp x, v$$

But this can be written as follows, with $J^\perp$ standing for the plane orthogonal to J:

$$x, v \in J^\perp$$

Now since J is fixed by Theorem 5.9, we conclude that both x, v, and in particular the position x, and so the whole trajectory, lie in this fixed plane $J^\perp$, as claimed.

(2) Conversely now, forget about Theorem 5.9, and assume that the trajectory lies in a certain plane E. Thus $x \in E$, and by differentiating we have $v \in E$ too, and so $x, v \in E$. Thus $E = J^\perp$, and so $J = E^\perp$, so the direction of J is fixed, as claimed.

(3) Regarding now the last assertion, we know from standard classical mechanics, as formulated by Newton, that the Kepler 2 law is more or less equivalent to the formula $\dot{\theta} = \lambda/r^2$. However, the derivation of $\dot{\theta} = \lambda/r^2$ was something tricky, and what we want to prove now is that this appears as a simple consequence of $||J|| = \text{constant}$.

(4) In order to to so, let us compute J, according to its definition $J = x \times p$, but in polar coordinates, which will change everything. Since $p = m\dot{x}$, we have:

$$J = r \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix} \times m \begin{pmatrix} \dot{r}\cos\theta - r\sin\theta \cdot \dot{\theta} \\ \dot{r}\sin\theta + r\cos\theta \cdot \dot{\theta} \\ 0 \end{pmatrix}$$

Now recall from the definition of the vector product that we have:

$$\begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \times \begin{pmatrix} c \\ d \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ ad - bc \end{pmatrix}$$

Thus J is a vector of the above form, with its last component being:

$$\begin{aligned} J_z &= rm\begin{vmatrix}\cos\theta & \dot{r}\cos\theta - r\sin\theta\cdot\dot{\theta}\\ \sin\theta & \dot{r}\sin\theta + r\cos\theta\cdot\dot{\theta}\end{vmatrix}\\ &= rm\cdot r(\cos^2\theta+\sin^2\theta)\dot{\theta}\\ &= r^2m\cdot\dot{\theta}\end{aligned}$$

(5) Now with the above formula in hand, our claim is that the magnitude $||J||$ is constant precisely when $\dot{\theta}=\lambda/r^2$, for some $\lambda\in\mathbb{R}$. Indeed, up to the obvious fact that the orientation of J is a binary parameter, who cannot just switch like that, let us just agree on this, knowing J is the same as knowing J_z, and is also the same as knowing $||J||$. Thus, our claim is proved, and this leads to the conclusion in the statement. □

As another basic application of the vector products, still staying with classical mechanics, we have all sorts of useful formulae regarding rotating frames. We first have:

Theorem 5.11. *Assume that a* 3*D body rotates along an axis, with angular speed w. For a fixed point of the body, with position vector x, the usual* 3*D speed is*

$$v=\omega\times x$$

where $\omega=wn$, with n unit vector pointing North. When the point moves on the body

$$V=\dot{x}+\omega\times x$$

is its speed computed by an inertial observer O on the rotation axis.

Proof. We have two assertions here, both requiring some 3D thinking, as follows:

(1) Assuming that the point is fixed, the magnitude of $\omega\times x$ is the good one, due to the following computation, with r being the distance from the point to the axis:

$$||\omega\times x||=w||x||\sin t=wr=||v||$$

As for the orientation of $\omega\times x$, this is the good one as well, because the North pole rule used above amounts in applying the right-hand rule for finding n, and so ω, and this right-hand rule was precisely the one used in defining the vector products $\times$.

(2) Next, when the point moves on the body, the inertial observer O can compute its speed by using a frame (u_1,u_2,u_3) which rotates with the body, as follows:

$$\begin{aligned} V &= \dot{x}_1u_1+\dot{x}_2u_2+\dot{x}_3u_3+x_1\dot{u}_1+x_2\dot{u}_2+x_3\dot{u}_3\\ &= \dot{x}+(x_1\cdot\omega\times u_1+x_2\cdot\omega\times u_2+x_3\cdot\omega\times u_3)\\ &= \dot{x}+w\times(x_1u_1+x_2u_2+x_3u_3)\\ &= \dot{x}+\omega\times x\end{aligned}$$

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

In what regards now the acceleration, the result, which is famous, is as follows:

THEOREM 5.12. *Assuming as before that a 3D body rotates along an axis, the acceleration of a moving point on the body, computed by O as before, is given by*

$$A = a + 2\omega \times v + \omega \times (\omega \times x)$$

with $\omega = wn$ being as before. In this formula the second term is called Coriolis acceleration, and the third term is called centripetal acceleration.

PROOF. This comes by using twice the formulae in Theorem 5.11, as follows:

$$\begin{aligned} A &= \dot{V} + \omega \times V \\ &= (\ddot{x} + \dot{\omega} \times x + \omega \times \dot{x}) + (\omega \times \dot{x} + \omega \times (\omega \times x)) \\ &= \ddot{x} + \omega \times \dot{x} + \omega \times \dot{x} + \omega \times (\omega \times x) \\ &= a + 2\omega \times v + \omega \times (\omega \times x) \end{aligned}$$

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

The truly famous result is actually the one regarding forces, obtained by multiplying everything by a mass m, and writing things the other way around, as follows:

$$ma = mA - 2m\omega \times v - m\omega \times (\omega \times x)$$

Here the second term is called Coriolis force, and the third term is called centrifugal force. These forces are both called apparent, or fictious, because they do not exist in the inertial frame, but they exist however in the non-inertial frame of reference, as explained above. And with of course the terms centrifugal and centripetal not to be messed up.

In fact, even more famous is the terrestrial application of all this, as follows:

THEOREM 5.13. *The acceleration of an object m subject to a force F is given by*

$$ma = F - mg - 2m\omega \times v - m\omega \times (\omega \times x)$$

with g pointing upwards, and with the last terms being the Coriolis and centrifugal forces.

PROOF. This follows indeed from the above discussion, by assuming that the acceleration A there comes from the combined effect of a force F, and of the usual g. □

We refer to any standard undergraduate mechanics book, such as Feynman [33], Kibble [57] or Taylor [90] for more on the above, including various numerics on what happens here on Earth, the Foucault pendulum, history of all this, and many other things. Let us just mention here, as a basic illustration for all this, that a rock dropped from 100m deviates about 1cm from its intended target, due to the formula in Theorem 5.13.

5c. Angular momentum

Back now to our quantum physics business, let us first talk about angular momentum, and we will get to spin later. We will need the following basic result:

Proposition 5.14. *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, $x_ix_j=x_jx_i$, 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 5.15. *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 5.14, we have:

$$[yp_z, zp_x] = yp_z zp_x - zp_x yp_z = y(zp_z - ih)p_x - zyp_x p_z = -ihyp_x$$
$$[yp_z, xp_z] = yp_z xp_z - xp_z yp_z = 0$$
$$[zp_y, zp_x] = zp_y zp_x - zp_x zp_y = 0$$
$$[zp_y, xp_z] = zp_y xp_z - xp_z zp_y = zxp_y p_z - x(zp_z - ih)p_y = ihxp_y$$

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 × used, and so can indeed be invoked. □

As an interesting consequence of Theorem 5.15, we have:

Proposition 5.16. *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 5.15. □

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 5.16, 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 5.17. *The operators L^2,L_x diagonalize as*

$$L^2f_l^m=h^2l(l+1)f_l^m$$

$$L_xf_l^m=hmf_l^m$$

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 $L^2f=\lambda f$, $L_xf=\mu f$ 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:

$$\begin{aligned}L^2(L_\pm f)&=L_\pm(L^2f)\\&=L_\pm(\lambda f)\\&=\lambda(L_\pm f)\end{aligned}$$

As for the second formula, this follows from:

$$\begin{aligned}L_x(L_\pm f)&=L_xL_\pm f\\&=(L_xL_\pm-L_\pm L_x)f+L_\pm L_xf\\&=\pm hL_\pm f+L_\pm(\mu f)\\&=(\mu\pm h)(L_\pm f)\end{aligned}$$

(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]. □

Summarizing, we have a good theoretical understanding of the angular momentum.

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 5.18. *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].

5d. Electron spin

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 5.19. *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 5.19 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 5.20. *We have the following diagonalization formulae*

$$S^2f_s^m=h^2s(s+1)f_s^m$$
$$S_xf_s^m=hmf_s^m$$
$$S_\pm f_s^m=h\sqrt{s(s+1)-m(m\pm1)}\,f_s^{m\pm1}$$

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:

$$S_\pm f_s^m=C_{s,\pm}^mf_s^{m\pm1}$$

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. $\square$

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 5.20 reformulate, in this $H=\mathbb C^2$ setting. But here, not many choices, and we are led to:

Definition 5.21. *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 $f=e_1,e_2$,

$$S^2f=h^2s(s+1)f$$
$$S_xf=hm_ff$$
$$S_\pm f=h\sqrt{s(s+1)-m_f(m_f\pm1)}\,\check{f}$$

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 5.20, because these are the choices providing a "minimal" realization of the equations in Theorem 5.20, in the smallest possible number of dimensions, namely $N=2$.

The point now is that the above questions can be solved, the result being:

Theorem 5.22. *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 5.21, 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 5.23. *Electrons have spin* 1/2.

This is motivated of course by the formula $s=1/2$ in Definition 5.21, 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 ±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.

5e. Exercises

Exercises:

Exercise 5.24.

Exercise 5.25.

Exercise 5.26.

Exercise 5.27.

Exercise 5.28.

Exercise 5.29.

Exercise 5.30.

Exercise 5.31.

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.

6b. Relativistic correction

Moving ahead now with corrections to Theorem 6.3, as a first goal, we would like to perform a relativistic correction. And for this purpose, we must go back to relativity theory, as discussed in the beginning of this book. So, as a continuation of the material there, let us discuss what happens to momentum, mass and energy. We would like to fix the momentum conservation equations for the plastic collisions, namely:

$$m=m_1+m_2$$

$$mv=m_1v_1+m_2v_2$$

However, this cannot really be done with bare hands, and by this meaning with mathematics only. But with some help from experiments, the conclusion is as follows:

Fact 6.4. *When defining the relativistic mass of an object of rest mass $m>0$, moving at speed v, by the formula*

$$M=\gamma m\quad:\quad\gamma=\frac{1}{\sqrt{1-v^2/c^2}}$$

this relativistic mass M, and the corresponding relativistic momentum $P=Mv$, are both conserved during collisions.

In other words, the situation here is a bit similar to that of the Galileo addition vs Einstein addition for speeds. The collision equations given above are in fact low-speed approximations of the correct, relativistic equations, which are as follows:

$$M = M_1 + M_2$$

$$Mv = M_1v_1 + M_2v_2$$

It remains now to discuss kinetic energy. You have certainly heard of the formula $E = mc^2$, which might actually well be on your T-shirt, now as you read this book, and in this case here is the explanation for it, in relation with the above:

Theorem 6.5. *The relativistic energy of an object of rest mass $m > 0$,*

$$\mathcal{E} = Mc^2 \quad : \quad M = \gamma m$$

which is conserved, as being a multiple of M, can be written as $\mathcal{E} = E + T$, with

$$E = mc^2$$

being its $v = 0$ component, called rest energy of m, and with

$$T = (1 - \gamma)mc^2 \simeq \frac{mv^2}{2}$$

being called relativistic kinetic energy of m.

Proof. All this is a bit abstract, coming from Fact 6.4, as follows:

(1) Given an object of rest mass $m > 0$, consider its relativistic mass $M = \gamma m$, as appearing in Fact 6.4, and then consider the following quantity:

$$\mathcal{E} = Mc^2$$

We know from Fact 6.4 that the relativistic mass M is conserved, so $\mathcal{E} = Mc^2$ is conserved too. In view of this, is makes somehow sense to call $\mathcal{E}$ energy. There is of course no clear reason for doing that, but let's just do it, and we'll understand later.

(2) Let us compute $\mathcal{E}$. This quantity is by definition given by:

$$\mathcal{E} = Mc^2 = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}}$$

Since $1/\sqrt{1-x} \simeq 1 + x/2$ for x small, by calculus, we obtain, for v small:

$$\mathcal{E} \simeq mc^2 \left(1 + \frac{v^2}{2c^2}\right) = mc^2 + \frac{mv^2}{2}$$

And, good news here, we recognize at right the kinetic energy of m.

(3) But this leads to the conclusions in the statement. Indeed, we are certainly dealing with some sort of energies here, and so calling the above quantity $\mathcal{E}$ relativistic energy

is legitimate, and calling $E = mc^2$ rest energy is legitimate too. Finally, the difference between these two energies $T = \mathcal{E} - E$ follows to be given by:

$$T = (1-\gamma)mc^2 \simeq \frac{mv^2}{2}$$

Thus, calling T relativistic kinetic energy is legitimate too, and we are done. $\square$

Now back to our business, hydrogen atom and quantum mechanics, inspired by the above, we have the following result, which is something non-trivial:

THEOREM 6.6. *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 the Einstein considerations above, coming from relativity, the relativistic kinetic energy is given by the following formula:

$$T = \frac{p^2}{2m} - \frac{p^4}{8m^3c^2} + \ldots$$

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 $\mathcal{E}_{nl}$. $\square$

6c. Magnetic moment

Equally non-trivial is the following correction, independent from the above one:

THEOREM 6.7. *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 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 $\mathcal{E}_{nj}$. $\square$

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 6.8. *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:

$$\begin{aligned}\mathcal{E}_{nl}+\mathcal{E}_{nj}&=\frac{\alpha^2E_n}{n^2}\left(\frac{n}{l+1/2}-\frac{3}{4}-\frac{n(j-l)}{(l+1/2)(j+1/2)}\right)\\&=\frac{\alpha^2E_n}{n^2}\left(\frac{n}{l+1/2}\left(1-\frac{j-l}{j+1/2}\right)-\frac{3}{4}\right)\\&=\frac{\alpha^2E_n}{n^2}\left(\frac{n}{j+1/2}-\frac{3}{4}\right)\end{aligned}$$

Thus the corrected formula of the energy is as follows:

$$\begin{aligned}E_{nj}&=E_n+\mathcal{E}_{nl}+\mathcal{E}_{nj}\\&=E_n+\frac{\alpha^2E_n}{n^2}\left(\frac{n}{j+1/2}-\frac{3}{4}\right)\end{aligned}$$

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 6.9. *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. □

6d. Further corrections

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.

6e. Exercises

Exercises:

Exercise 6.10.

Exercise 6.11.

Exercise 6.12.

Exercise 6.13.

Exercise 6.14.

Exercise 6.15.

Exercise 6.16.

Exercise 6.17.

Bonus exercise.

CHAPTER 7

Dirac equation

7a. Klein-Gordon

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:

$$\begin{aligned} x' &= \gamma(x - vt) \\ y' &= y \\ z' &= z \\ t' &= \gamma(t - vx/c^2) \end{aligned}$$

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.

7b. Dirac equation

Getting to work now, following Dirac, the idea will be that 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.

7c. Plane waves

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,

$$h\big(\gamma_0f-\gamma_1g_1-\gamma_2g_2-\gamma_3g_3\big)u=mcu$$

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,

$$v=\frac{<g,\sigma>}{f-mc/h}w\quad,\quad w=\frac{<g,\sigma>}{f+mc/h}v$$

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:

$$\begin{aligned}&\Big(h\big(\gamma_0f-\gamma_1g_1-\gamma_2g_2-\gamma_3g_3\big)-mc\Big)u\\ =\ &\begin{pmatrix}hf-mc&-h<g,\sigma>\\ h<g,\sigma>&-hf-mc\end{pmatrix}\begin{pmatrix}v\\ w\end{pmatrix}\\ =\ &\begin{pmatrix}(hf-mc)v-h<g,\sigma>w\\ h<g,\sigma>v-(hf+mc)w\end{pmatrix}\end{aligned}$$

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*

$$||g||^2=f^2-\left(\frac{mc}{h}\right)^2$$

under the assumption that the solution is nonzero, $u\neq 0$.

Proof. Consider the equations found in Proposition 7.14, namely:

$$v=\frac{<g,\sigma>}{f-mc/h}w\quad,\quad w=\frac{<g,\sigma>}{f+mc/h}v$$

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.

7d. Invariance questions

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 $\gamma=1/\sqrt{1-\beta^2}$:

$$ct'=\gamma(ct-\beta x)$$

$$x'=\gamma(x-\beta ct)$$

$$y'=y$$

$$z'=z$$

Equivalently, in matrix form, we have the following formula:

$$\begin{pmatrix}ct'\\x'\\y'\\z'\end{pmatrix}=\begin{pmatrix}\gamma&-\gamma\beta&0&0\\-\gamma\beta&\gamma&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\begin{pmatrix}ct\\x\\y\\z\end{pmatrix}$$

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:

$$A\gamma_0A=\gamma_0$$

$$A^{-1}\gamma_1A=\gamma_1$$

$$A\gamma_2A=\gamma_2$$

$$A^{-1}\gamma_3A=\gamma_3$$

(9) Indeed, the first formula comes from the following computation:

$$\begin{aligned}A\gamma_0A &= (a+b\gamma_0\gamma_2)\gamma_0(a+b\gamma_0\gamma_2)\\ &= a^2\gamma_0+ab\gamma_0\gamma_0\gamma_2+ab\gamma_0\gamma_2\gamma_0+b^2\gamma_0\gamma_2\gamma_0\gamma_0\gamma_2\\ &= a^2\gamma_0-b^2\gamma_0\\ &= \gamma_0\end{aligned}$$

The second formula comes from a similar computation, as follows:

$$\begin{aligned}A^{-1}\gamma_1A &= (a+b\gamma_0\gamma_2)\gamma_1(a-b\gamma_0\gamma_2)\\ &= a^2\gamma_1-ab\gamma_1\gamma_0\gamma_2+ab\gamma_0\gamma_2\gamma_1-b^2\gamma_0\gamma_2\gamma_1\gamma_0\gamma_2\\ &= a^2\gamma_1-b^2\gamma_1\\ &= \gamma_1\end{aligned}$$

The third formula again comes from a similar computation, as follows:

$$\begin{aligned}A\gamma_2A &= (a+b\gamma_0\gamma_2)\gamma_2(a+b\gamma_0\gamma_2)\\ &= a^2\gamma_2+ab\gamma_2\gamma_0\gamma_2+ab\gamma_0\gamma_2\gamma_2+b^2\gamma_0\gamma_2\gamma_2\gamma_0\gamma_2\\ &= a^2\gamma_2-b^2\gamma_2\\ &= \gamma_2\end{aligned}$$

As for the fourth formula, this comes again from a similar computation, namely:

$$\begin{aligned}A^{-1}\gamma_3A &= (a+b\gamma_0\gamma_2)\gamma_3(a-b\gamma_0\gamma_2)\\ &= a^2\gamma_3-ab\gamma_3\gamma_0\gamma_2+ab\gamma_0\gamma_2\gamma_3-b^2\gamma_0\gamma_2\gamma_3\gamma_0\gamma_2\\ &= a^2\gamma_3-b^2\gamma_3\\ &= \gamma_3\end{aligned}$$

(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:

$$\begin{aligned}A^{-1}\gamma_0A &= (a-b\gamma_0\gamma_2)\gamma_0(a+b\gamma_0\gamma_2)\\ &= a^2\gamma_0+ab\gamma_0\gamma_0\gamma_2-ab\gamma_0\gamma_2\gamma_0-b^2\gamma_0\gamma_2\gamma_0\gamma_0\gamma_2\\ &= (a^2+b^2)\gamma_0+2ab\gamma_2\\ &= \gamma\cdot\gamma_0-\gamma\beta\cdot\gamma_2\end{aligned}$$

Similarly, we have the following computation:

$$\begin{aligned}A^{-1}\gamma_2A &= (a-b\gamma_0\gamma_2)\gamma_2(a+b\gamma_0\gamma_2)\\ &= a^2\gamma_2+ab\gamma_2\gamma_0\gamma_2-ab\gamma_0\gamma_2\gamma_2-b^2\gamma_0\gamma_2\gamma_2\gamma_0\gamma_2\\ &= (a^2+b^2)\gamma_2-2ab\gamma_0\\ &= \gamma\cdot\gamma_2+\gamma\beta\cdot\gamma_0\end{aligned}$$

(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 $A^{-1}\gamma_iA=\gamma_iL$. But because of $\gamma_i^2=1$, we can reformulate them in the following way:

$$L=\gamma_iA^{-1}\gamma_iA$$

Now in view of the above, it makes sense to introduce the following matrices:

$$L_i=\gamma_iA^{-1}\gamma_iA$$

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. ☐

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.

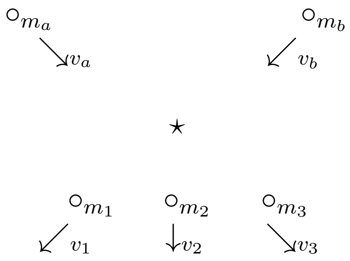
CHAPTER 8

Feynman diagrams

8a. Decay, scattering

We have seen in the previous chapter that, with Dirac and others helping, it is relatively easy to escape from the electron/photon world we've been living in, so far in this book, with the introduction of other particles, appearing from equations inspired from the Schrödinger equation for the electron, and the wave equation for the photon.

In order to further advance, as to be ready to come back to hydrogen, we need to talk about interactions between our particles. But here, we already have some experience from classical mechanics, with the typical picture of what can happen being as follows:



To be more precise, this diagram describes a collision between two particles, but we can of course allow further particles entering the collision, and then the several particles emerging from this collision. The data of each particle, which in classical mechanics means mass m and speed v, is carefully recorded, with of course the aim of recovering the output data from the input data. Finally, all this can take place in arbitrary dimensions, and also, importantly, the collision can be elastic, or plastic, or any mixture of these.

This was for basic interactions in classical mechanics. In our present setting, particle physics, things are a bit more complicated than this, due to a variety of reasons, and experimental physics suggests looking at two main types of interactions, as follows:

Fact 8.1. *In particle physics, we have two main types of interactions, namely:*

1. *Decay. This is when a particle decomposes, as a result of whatever internal mechanism, into a sum of other particles,* $*_0 \to *_1 + \ldots + *_n$.
2. *Scattering. This is when two particles meet, by colliding, or almost, and combine and decompose into a sum of other particles,* $*_a + *_b \to *_1 + \ldots + *_n$.

Obviously, all this departs a bit from our classical mechanics knowledge, as explained above, and several comments are in order here, as follows:

(1) In what regards decay, something that we talked a lot about, when doing thermodynamics, and then quantum mechanics, is an electron of an atom changing its energy level, and emitting a photon. But this can be regarded as being decay.

(2) As for scattering, the simplest example here appears again from an electron of an atom, changing its energy level, but this time by absorbing a photon. Of course, there are many other possible examples, such as the electron-positron annihilation.

(3) Regarding now the mechanisms at work, for decay we certainly have some intuition from classical mechanics, and we can label that process as being some sort of "explosion". With of course the comment that usual explosions are rather something chemical.

(4) As for scattering, this normally stands for some sort of "collision", a bit as in classical mechanics, but with the comment that we are really talking here about general scattering, not only collisions. More on this, which is something quite subtle, later.

Getting to work for good now, let us first gather some knowledge and data about decay. However, this is no easy business, in view of the physics that we know, and it is helpful at this point to take off, and get into our popular physics knowledge, regarding radioactivity. Or at least that is easy for me, during Chernobyl 1986 I used to be a teenager in the nearby Romania, and we all duly learned, in a hurry, all that theory. But you surely know a bit about this too, say from Fukushima 2011, don't you.

So, decay and its mathematics. Ignoring the physics, this is basically a matter of probability and statistics, and the basics here can be summarized as follows:

Theorem 8.2. *In the context of decay, the quantity to look at is the decay rate λ, which is the probability per unit time that the particle will disintegrate. With this:*

1. *The number of particles remaining at time $t > 0$ is $N_t = e^{-\lambda t} N_0$.*
2. *The mean lifetime of a particle is $\tau = 1/\lambda$.*
3. *The half-life of the substance is $t_{1/2} = (\log 2)/\lambda$.*

Proof. As said above, this is basic probability, as follows:

(1) In mathematical terms, our definition of the decay rate reads:

$$\frac{dN}{dt} = -\lambda N$$

By integrating, we are led to the formula in the statement, namely:

$$N_t = e^{-\lambda t} N_0$$

(2) Let us first convert what we have into a probability law. We have:

$$\int_0^\infty N_t dt = \int_0^\infty N_0 e^{-\lambda t} dt = \frac{N_0}{\lambda}$$

Thus, the density of the probability decay function is given by:

$$f(t) = \frac{\lambda}{N_0} \cdot N_0 e^{-\lambda t} = \lambda e^{-\lambda t}$$

We can now compute the mean lifetime, by integrating by parts, as follows:

$$\begin{aligned}
\tau &= < t > \\
&= \int_0^\infty t f(t) dt \\
&= \int_0^\infty \lambda t e^{-\lambda t} dt \\
&= \int_0^\infty t(-e^{-\lambda t})' dt \\
&= \int_0^\infty e^{-\lambda t} dt \\
&= \frac{1}{\lambda}
\end{aligned}$$

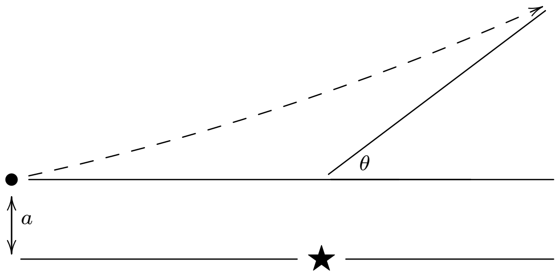
(3) Finally, regarding the half-life, this is by definition the time $t_{1/2}$ required for the decaying quantity to fall to one-half of its initial value. Mathematically, this means:

$$N_t = 2^{-\frac{t}{t_{1/2}}} N_0$$

Now by comparing with $N_t = e^{-\lambda t} N_0$, this gives $t_{1/2} = (\log 2)/\lambda$, as stated. $\square$

Getting now to scattering, this is something far more familiar, because we can fully use here our experience from classical mechanics. Let us start with:

DEFINITION 8.3. *The generic picture of scattering is as follows,*



with $a \geq 0$ being the impact parameter, and $\theta \in [0, \pi]$ being the scattering angle.

In other words, we assume here that the particle misses its target by $a \geq 0$, with the limiting case $a = 0$ corresponding of course to exactly hitting the target, and we are interested in computing the scattering angle $\theta \in [0, \pi]$ as a function $\theta = \theta(a)$.

As an important comment now, we are interested in $a > 0$, because this is what happens in particle physics, there is no need for exactly hitting the target for having a collision-type interaction. By the case, the limiting case $a = 0$ is rather unwanted in the context of our scaterring question, because by symmetry this would normally force the scattering angle to be $\theta = 0$ or $\theta = \pi$, which does not look very interesting.

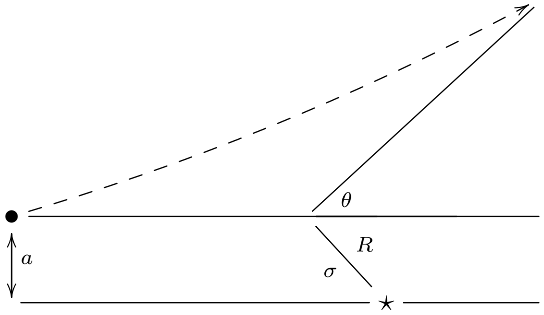
But probably too much talking, let us do a computation. We have here:

Proposition 8.4. *In the context of classical particle colliding elastically with a hard sphere of radius $R > 0$, we have the formula*

$$a = R \cos \frac{\theta}{2}$$

and so the scattering angle is given by $\theta = 2 \arccos(a/R)$.

Proof. In the context from the statement, which is all classical mechanics, and more specifically is a basic elastic collision, between a point particle and a hard sphere, if the impact factor is $a > R$, nothing happens. In the case $a \leq R$ we do have an impact, and a bounce of our particle on the hard sphere, the picture of the event being as follows:



Here the sphere is missing, due to budget cuts, with only its center $\star$ being pictured, but you get the point. Now with σ being the angle in the statement, we have the following two formulae, with the first one being clear on the above picture, and with the second one coming from the fact that, at the rebound, the various angles must sum up to π:

$$a = R \sin \sigma \quad , \quad 2\sigma + \theta = \pi$$

We deduce that the impact factor is given by the following formula:

$$a = R \sin \left(\frac{\pi}{2} - \frac{\theta}{2} \right) = R \cos \frac{\theta}{2}$$

Thus, we are led to the conclusions in the statement. $\square$

With this understood, let us try to make something more 3D, and statistical, out of this. We can indeed further build on Definition 8.3, as follows:

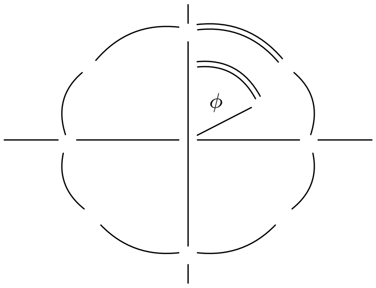
Definition 8.5. *In the general context of scattering, we can:*

1. *Extend our length/angle correspondence $a \to \theta$ into an infinitesimal area/solid angle correspondence $d\sigma \to d\Omega$.*
2. *Talk about the inverse derivative $D(\theta)$ of this correspondence, called differential cross section, according to the formula $d\sigma = D(\theta)d\Omega$.*
3. *And finally, define the total cross section of the scattering event as being the quantity $\sigma = \int d\sigma = \int D(\theta)d\Omega$.*

And good news, the notion of total cross section σ, as constructed above, is the one that we will need, in what follows, with this being to scattering something a bit similar to what the decay rate λ was to decay, that is, the main quantity to look at.

In order to understand how the cross section works, we have:

Proposition 8.6. *Assuming that the incoming beam comes as follows,*



subtending a certain angle ϕ, the differential cross section is given by

$$D(\theta) = \left|\frac{a}{\sin\theta} \cdot \frac{da}{d\theta}\right|$$

and the total cross section is given by $\sigma = \int D(\theta)d\Omega$.

Proof. Assume indeed that we have a uniform beam as the one pictured in the statement, enclosed by the double lines appearing there, and with the need for a beam instead of a single particle coming from what we do in Definition 8.5, which is rather of continuous nature. Our claim is that we have the following formulae:

$$d\sigma = |a \cdot da \cdot d\phi| \quad , \quad d\Omega = |\sin\theta \cdot d\theta \cdot d\phi|$$

Indeed, the first formula, at departure, is clear from the picture above, and the second formula is clear from a similar picture at the arrival. Now with these formulae in hand, by dividing them, we obtain the following formula for the differential cross section:

$$\begin{aligned}D(\theta) &= \frac{d\sigma}{d\Omega}\\ &= \left|\frac{a\cdot da\cdot d\phi}{\sin\theta\cdot d\theta\cdot d\phi}\right|\\ &= \left|\frac{a}{\sin\theta}\cdot\frac{da}{d\theta}\right|\end{aligned}$$

As for the total cross section, this is given as usual by $\sigma=\int D(\theta)d\Omega$. □

As an illustration for this, in the case of a hard sphere scattering, we have:

THEOREM 8.7. *In the case of a hard sphere scattering, the cross section is*

$$\sigma=\pi R^2$$

with $R>0$ being the radius of the sphere.

PROOF. We know from Proposition 8.4 that, with the notations there, we have:

$$a=R\cos\frac{\theta}{2}$$

At the level of the corresponding differentials, this gives the following formula:

$$\frac{da}{d\theta}=-\frac{R}{2}\sin\frac{\theta}{2}$$

We can now compute the differential cross section, as above, and we obtain:

$$\begin{aligned}D(\theta) &= \left|\frac{a}{\sin\theta}\cdot\frac{da}{d\theta}\right|\\ &= \frac{R\cos(\theta/2)}{\sin\theta}\cdot\frac{R\sin(\theta/2)}{2}\\ &= \frac{R^2(\sin\theta)/2}{2\sin\theta}\\ &= \frac{R^2}{4}\end{aligned}$$

Now by integrating, we obtain from this, via some calculus, the following formula:

$$\sigma=\int\frac{R^2}{4}\,d\Omega=\pi R^2$$

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

Summarizing, for a hard sphere scattering, the cross section turns to be something very simple, namely the area of the sphere met by the beam. This is of course something quite particular, and when using more complicated targets, the formula of σ gets more complicated too. We will come back to this, with further examples, later on.

8b. Golden Rule

Time now to get into the real thing, namely quantum electrodynamics (QED) and Feynman diagrams, which correctly explain the behavior of the electron, and of other particles that we know. And to be followed later by quantum chromodynamics (QCD), which does even better, dealing with smaller beasts, quarks and related gnomes.

We will be following for our presentation the lovely particle physics book of Griffiths [44]. That is a remarkable book, taking the challenge of explaining such things, which are normally quite advanced, to undergraduates. Of course, in case you are not exactly an undergraduate, say you are a professional mathematical physicist, and struggle with what comes next, do not worry. This is normal, can happen to everyone, and I must admit that I struggled too, sometimes quite late in my career, with such things.

Skipping a discussion here with cat, whose sardonic smile, when hearing all this, does not look very inviting, let us get to work. As a main principle regarding particle decay, following Fermi and others, we have the following simple and useful fact:

Principle 8.8 (Fermi Golden Rule). *In the context of a particle physics decay,* $*_0 \to *_1 + \ldots + *_n$, *the decay rate is given by*

$$\lambda = \int |M|^2 dp$$

with $M = M(p_0, \ldots, p_n)$ *being the amplitude of the interaction, and with the integration being restricted to the part of the phase space allowed by basic physics.*

Obviously, several things going on here, that will take us some time, to understand. To start with, the above Golden Rule looks quite reasonable, namely getting λ by integrating something on the phase space. It remains to understand two things, namely what the formula of the amplitude M is, and where does the integration exactly take place.

Leaving the formula of the amplitude M for later, let us try to answer the second question, regarding the allowed phase space. According to the Golden Rule, that is simply the phase space allowed by basic physics, and here that basic physics is:

Addendum 8.9. *In the above context, the basic physics is as follows:*

1. *The total energy and momentum must be conserved.*
2. *Each outgoing particle must keep its mass constant.*
3. *Each outgoing particle must have positive energy.*

Summarizing, all common sense things that we have here. In mathematical terms now, it is better to add integrands corresponding to the above conditions (1,2,3), instead of exactly specifying the allowed state space. And with both (1) and (2) requiring Dirac masses δ, and with (3) requiring a Heaviside function $H = \chi_{(0,\infty)}$, we are led to:

Principle 8.10 (Golden Rule 2). *In the context of a particle physics decay, $*_0 \to *_1 + \ldots + *_n$, the decay rate is given by*

$$\lambda = \int |M|^2 \delta\left(p_0 - \sum_{i=1}^n p_i\right) \prod_{i=1}^n \delta(p_i^2 - m_i^2 c^2) H(p_i^0) dp$$

with $M = M(p_0, \ldots, p_n)$ being the amplitude of the interaction.

Which looks quite neat, but there is actually a subtlety here, in relation with the Dirac masses, which take as arguments squares of variables, instead of the variables themselves. In order to clarify this, let us make the following computation, with $a > 0$:

$$\begin{aligned}
\int_{\mathbb{R}} f(x)\delta(x^2 - a^2)dx &= \int_{-\infty}^0 f(x)\delta(x^2 - a^2)dx + \int_0^\infty f(x)\delta(x^2 - a^2)dx \\
&= \int_{-\infty}^a f(y-a)\delta(y^2 - 2ay)dy + \int_{-a}^\infty f(y+a)\delta(y^2 + 2ay)dy \\
&\simeq \int_{-\infty}^a f(y-a)\delta(-2ay)dy + \int_{-a}^\infty f(y+a)\delta(2ay)dy \\
&= \int_{-\infty}^{2a^2} f\left(\frac{z}{2a} - a\right)\delta(-z)\frac{dz}{2a} + \int_{-2a^2}^\infty f\left(\frac{z}{2a} + a\right)\delta(z)\frac{dz}{2a} \\
&= \frac{f(-a)}{2a} + \frac{f(a)}{2a} \\
&= \int_R f(x)\,\frac{\delta(x-a) + \delta(x+a)}{2a}\,dx
\end{aligned}$$

Sounds like physics, you would say, and in answer, yes physics that is, but in any case, we have in this way the definition for our quadratic Dirac masses, as follows:

$$\delta(x^2 - a^2) = \frac{\delta(x-a) + \delta(x+a)}{2a}$$

With this understood, and before getting into what the amplitude M is, let us make some normalizations. Here these normalizations are, and you will have to believe me here, all of them are made for good reasons, as we will discover in a moment:

(1) We have $\lambda \sim S$, with $S = 1/\prod_i (m_i!)$, where $m_i \in \mathbb{N}$ with $\sum m_i = n$ are the multiplicities of the output particles, and it is better to leave S outside the integral.

(2) Also, $\lambda \sim 1/(2hm_0)$, with h being as usual the reduced Planck constant, and m_0 being the initial mass, and it is better to leave $1/(2hm_0)$ outside the integral too.

(3) Each Dirac mass δ behaves better in computations when multiplied by a 2π factor. Also, each individual dp_i symbol behaves better when divided by a 2π factor.

Now by doing all these normalizations, which amounts in correspondingly rescaling the amplitude M, and with this being certainly not a big deal, because we haven't even talked yet about what this amplitude M is, so free to do this, we are led to:

PRINCIPLE 8.11 (Golden Rule 3). *In the context of a particle physics decay, $*_0 \to *_1 + \ldots + *_n$, the decay rate is given by*

$$\lambda = \frac{S}{2hm_0}\int |\mathcal{M}|^2(2\pi)^4\delta\left(p_0 - \sum_{i=1}^n p_i\right)\prod_{i=1}^n 2\pi\delta(p_i^2 - m_i^2c^2)H(p_i^0)\frac{dp_i}{(2\pi)^4}$$

with $\mathcal{M} = \mathcal{M}(p_0, \ldots, p_n)$ being the normalized amplitude of the interaction, and with $S = 1/\prod_i(m_i!)$, where $m_i \in \mathbb{N}$ with $\sum m_i = n$ are the multiplicities of the output.

And good news, this will be normally the final form of the Golden Rule for decays, that we will be using, in what follows. In practice, however, we will see in a moment that the integration with respect to time is easy to perform, and this will lead to yet another formulation of the Golden Rule, which is the most useful one, for applications.

Before that, however, some philosophical comments. The Golden Rule has become now something quite complicated, and there is still a discussion about $\mathcal{M}$, which will certainly bring its part of complicated mathematics. But remember that, in the end, everything comes from Principle 8.8, which is something quite simple. So, no fear.

This being said, even when looking at Principle 8.8, you might wonder, is that really correct, and where that really comes from. In answer, common sense as explained above, then lots of experiments too, confirming it, or rather confirming the formula of $\mathcal{M}$, that we haven't talked about yet, and finally quantum field theory, which is something advanced, that can actually prove this Golden Rule, starting from simple principles.

Back to our business now, we will take Principle 8.11 for granted, and further build on it, with examples, the formula of $\mathcal{M}$, and more. Before that, however, let us do what was suggested above, namely integrating with respect to time. This leads to:

THEOREM 8.12 (Golden Rule 4). *In the context of a particle physics decay, $*_0 \to *_1 + \ldots + *_n$, the decay rate is given, in standard $\tilde{p} = (E/c, p)$ notation, by*

$$\lambda = \frac{S}{2hm_0}\int |\mathcal{M}|^2(2\pi)^4\delta\left(\tilde{p}_0 - \sum_{i=1}^n \tilde{p}_i\right)\prod_{i=1}^n \frac{1}{2\sqrt{||p_i||^2 + m_i^2c^2}}\cdot\frac{dp_i}{(2\pi)^3}$$

with $\mathcal{M} = \mathcal{M}(p_0, \ldots, p_n)$ being the normalized amplitude, $S = 1/\prod_i(m_i!)$ being the statistical factor, and with the convention $E_i/c = \sqrt{||p_i||^2 + m_i^2c^2}$, both in $\mathcal{M}$ and δ.

PROOF. We use the formula from Principle 8.11, written with our standard notation for energy-momentum vectors $\tilde{p}=(E/c,p)$ from chapter 7, which is as follows:

$$\lambda=\frac{S}{2hm_0}\int|\mathcal{M}|^2(2\pi)^4\delta\left(\tilde{p}_0-\sum_{i=1}^n\tilde{p}_i\right)\prod_{i=1}^n 2\pi\delta(\tilde{p}_i^2-m_i^2c^2)H\left(\frac{E_i}{c}\right)\frac{d\tilde{p}_i}{(2\pi)^4}$$

In order to further process this formula, let us look at each of the n products on the right. According to our conventions for quadratic Dirac masses, explained after Principle 8.10, the Dirac mass appearing there is given by the following formula:

$$\begin{aligned}\delta(\tilde{p}_i^2-m_i^2c^2) &= \delta\left(\frac{E_i^2}{c^2}-||p_i||^2-m_i^2c^2\right)\\ &= \delta\left(\left(\frac{E_i}{c}\right)^2-\left(\sqrt{||p_i||^2+m_i^2c^2}\right)^2\right)\\ &= \frac{\delta\left(\frac{E_i}{c}-\sqrt{||p_i||^2+m_i^2c^2}\right)+\delta\left(\frac{E_i}{c}+\sqrt{||p_i||^2+m_i^2c^2}\right)}{2\sqrt{||p_i||^2+m_i^2c^2}}\end{aligned}$$

Thus we have two possibilities, and since the Heaviside term $H(E_i/c)$ equals 1 on the first one, and vanishes on the second one, we are led to the following formula:

$$\delta(\tilde{p}_i^2-m_i^2c^2)H\left(\frac{E_i}{c}\right)=\frac{\delta\left(\frac{E_i}{c}-\sqrt{||p_i||^2+m_i^2c^2}\right)}{2\sqrt{||p_i||^2+m_i^2c^2}}$$

But this leads to the conclusion in the statement. ☐

As an illustration, for two-particle decays many things simplify, and we have:

THEOREM 8.13. *For two-particle decays, $*_0\to *_1+*_2$, the Golden Rule reads*

$$\lambda=\frac{S||p||}{8\pi hm_0^2c}|\mathcal{M}|^2$$

with $\mathcal{M}$ being the amplitude, $||p||$ being the magnitude of either outgoing momentum,

$$||p||=\frac{c}{2m_0}\sqrt{m_0^4+m_1^4+m_2^4-2m_0^2m_1^2-2m_0^2m_2^2-2m_1^2m_2^2}$$

and the statistical factor being $S=1$ if $*_1\neq *_2$, and $S=1/2$ if $*_1=*_2$.

PROOF. In the case of two-particle decays, the formula in Theorem 8.12 takes the following form, with the statistical factor S being the one in the statement:

$$\begin{aligned}\lambda &= \frac{S}{2hm_0}\int|\mathcal{M}|^2(2\pi)^4\delta\left(\tilde{p}_0-\tilde{p}_1-\tilde{p}_2\right)\prod_{i=1}^2\frac{1}{2\sqrt{||p_i||^2+m_i^2c^2}}\cdot\frac{dp_i}{(2\pi)^3}\\ &= \frac{S}{32\pi^2hm_0}\int|\mathcal{M}|^2\frac{\delta\left(\tilde{p}_0-\tilde{p}_1-\tilde{p}_2\right)}{\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_2||^2+m_2^2c^2}}\,dp_1dp_2\end{aligned}$$

Let us look now at the Dirac function. This decomposes over components, as follows:

$$\delta\left(\tilde{p}_0-\tilde{p}_1-\tilde{p}_2\right)=\delta\left(\frac{E_0}{c}-\frac{E_1}{c}-\frac{E_2}{c}\right)\delta\left(p_0-p_1-p_2\right)$$

With the particle $*_0$ being supposed to be at rest, we have the following formulae:

$$\frac{E_0}{c}=m_0c\quad,\quad p_0=0$$

On the other hand, recall from Theorem 8.12 that the machinery there leads to:

$$\frac{E_1}{c}=\sqrt{||p_1||^2+m_1^2c^2}\quad,\quad\frac{E_2}{c}=\sqrt{||p_2||^2+m_2^2c^2}$$

Thus, the above Dirac mass is in fact given by the following formula:

$$\delta\left(\tilde{p}_0-\tilde{p}_1-\tilde{p}_2\right)=\delta\left(m_0c-\sqrt{||p_1||^2+m_1^2c^2}-\sqrt{||p_2||^2+m_2^2c^2}\right)\delta(p_1+p_2)$$

Getting back now to the formula of the decay rate, that becomes:

$$\begin{aligned}\lambda &= \frac{S}{32\pi^2hm_0}\int|\mathcal{M}|^2\frac{\delta\left(m_0c-\sqrt{||p_1||^2+m_1^2c^2}-\sqrt{||p_2||^2+m_2^2c^2}\right)}{\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_2||^2+m_2^2c^2}}\\ &\quad\times\delta(p_1+p_2)\,dp_1dp_2\end{aligned}$$

Since we must have $p_2=-p_1$, this expression further simplifies to:

$$\lambda=\frac{S}{32\pi^2hm_0}\int|\mathcal{M}|^2\frac{\delta\left(m_0c-\sqrt{||p_1||^2+m_1^2c^2}-\sqrt{||p_1||^2+m_2^2c^2}\right)}{\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_1||^2+m_2^2c^2}}\,dp_1$$

In spherical coordinates, this expression takes the following form:

$$\lambda=\frac{S}{32\pi^2hm_0}\int|\mathcal{M}|^2\frac{\delta\left(m_0c-\sqrt{r^2+m_1^2c^2}-\sqrt{r^2+m_2^2c^2}\right)}{\sqrt{r^2+m_1^2c^2}\sqrt{r^2+m_2^2c^2}}\,r^2\sin s\,drdsdt$$

The point now is that by physics, to be explained later, the amplitude must be of the form $\mathcal{M}=\mathcal{M}(r)$. Thus the angular integrals contribute with factors as follows:

$$\int_0^\pi\sin s\,ds=2\quad,\quad\int_0^{2t}dt=2\pi$$

We conclude that in the end we are left with a real integral, over r, as follows:

$$\lambda=\frac{S}{8\pi hm_0}\int_0^\infty|\mathcal{M}|^2\frac{\delta\left(m_0c-\sqrt{r^2+m_1^2c^2}-\sqrt{r^2+m_2^2c^2}\right)}{\sqrt{r^2+m_1^2c^2}\sqrt{r^2+m_2^2c^2}}\,r^2dr$$

In order to compute this integral, consider the following variable:

$$u=\sqrt{r^2+m_1^2c^2}+\sqrt{r^2+m_2^2c^2}$$

Now observe that by differentiating, we obtain the following formula:

$$\begin{aligned}\frac{du}{dr}&=\frac{2r}{2\sqrt{r^2+m_1^2c^2}}+\frac{2r}{2\sqrt{r^2+m_2^2c^2}}\\&=r\left(\frac{1}{\sqrt{r^2+m_1^2c^2}}+\frac{1}{\sqrt{r^2+m_2^2c^2}}\right)\\&=\frac{ur}{\sqrt{r^2+m_1^2c^2}\sqrt{r^2+m_2^2c^2}}\end{aligned}$$

Thus, in terms of this new variable u, we have the following formula:

$$\begin{aligned}\lambda&=\frac{S}{8\pi hm_0}\int_{m_1c+m_2c}^\infty|\mathcal{M}|^2\delta(m_0c-u)\frac{r}{u}\,du\\&=\frac{Sr}{8\pi hm_0^2c}\,|\mathcal{M}|^2\end{aligned}$$

Here in the last formula r stands for the value of the variable r evaluated at the place where the Dirac mass takes the value 1, that we can compute as follows:

$$\begin{aligned}u=m_0c\quad&\iff\quad\sqrt{r^2+m_1^2c^2}+\sqrt{r^2+m_2^2c^2}=m_0c\\&\iff\quad r^2+m_1^2c^2=r^2+m_2^2c^2+m_0^2c^2-2m_0c\sqrt{r^2+m_2^2c^2}\\&\iff\quad 2m_0\sqrt{r^2+m_2^2c^2}=(m_0^2-m_1^2+m_2^2)c\\&\iff\quad 4m_0^2(r^2+m_2^2c^2)=(m_0^2-m_1^2+m_2^2)^2c^2\\&\iff\quad 4m_0^2r^2=((m_0^2-m_1^2+m_2^2)^2-4m_0^2m_2^2)c^2\\&\iff\quad r=\frac{c}{2m_0}\sqrt{m_0^4+m_1^4+m_2^4-2m_0^2m_1^2-2m_0^2m_2^2-2m_1^2m_2^2}\end{aligned}$$

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

So long for particle decays, $*_0\to *_1+\ldots+*_n$. We will be back to them later in this chapter, once we will know more about the amplitude $\mathcal{M}$.

8c. Cross sections

With the particle decays understood, let us turn now to the other phenomenon that can appear, namely collision, or scattering, $*_a + *_b \to *_1 + \ldots + *_n$. Here, as you can imagine, the situation is quite similar to that for the decays, as follows:

(1) The Golden Rule in its first formulation, Principle 8.8, holds again, this time with a new amplitude for the interaction, of the form $N = N(p_a, p_b, p_1, \ldots .p_n)$, which remains, as before in the case of the decays, to be suitably normalized.

(2) The basic physics from Addendum 8.8, which complements Principle 8.8, namely conservation of total energy and total momentum, and constant mass and positive energy requirement for each outgoing particle, holds unchanged.

Thus, we led to the following analogue of Principle 8.10, or Golden Rule 2:

Principle 8.14 (Golden Rule 2'). *In the context of a particle physics scattering,* $*_a + *_b \to *_1 + \ldots + *_n$, *the cross section is given by*

$$\sigma = \int |N|^2 \delta\left(p_a + p_b - \sum_{i=1}^n p_i\right) \prod_{i=1}^n \delta(p_i^2 - m_i^2 c^2) H(p_i^0) dp$$

with $N = N(p_a, p_b, p_1, \ldots, p_n)$ *being the amplitude of the interaction.*

Regarding now the normalization of the amplitude, which is a key point, some things change at the level of physics, the situation being as follows:

(1) We have $\lambda \sim S$, with $S = 1/\prod_i (m_i!)$, where $m_i \in \mathbb{N}$ with $\sum m_i = n$ are the multiplicities of the output particles, exactly as before.

(2) Next, and importantly, at the level of the physics we have the following new formula, that we will discuss more in detail later, instead of the previous $\lambda \sim 1/(2hm_0)$, because we are doing now something else, with two particles colliding instead of one decaying, and by computing a cross section σ instead of a decay rate λ:

$$\lambda \sim \frac{h^2}{4\sqrt{< p_a, p_b >^2 - (m_a m_b c^2)^2}}$$

(3) Finally, as before, each Dirac mass δ behaves better in computations when multiplied by a 2π factor, and each dp_i behaves better when divided by a 2π factor.

Now by doing all these normalizations, and with some amplitude magic helping, we are led to the following analogue of Principle 8.11, or Golden Rule 3:

Principle 8.15 (Golden Rule 3'). *In the context of a particle physics scattering, $*_a + *_b \to *_1 + \ldots + *_n$, the cross section is given by*

$$\begin{aligned}\sigma &= \frac{Sh^2}{4\sqrt{<p_a,p_b>^2-(m_am_bc^2)^2}}\int|\mathcal{M}|^2(2\pi)^4\delta\left(p_a+p_b-\sum_{i=1}^np_i\right)\\ &\quad\times\prod_{i=1}^n2\pi\delta(p_i^2-m_i^2c^2)H(p_i^0)\frac{dp_i}{(2\pi)^4}\end{aligned}$$

with $\mathcal{M}=\mathcal{M}(p_a,p_b,p_1,\ldots,p_n)$ being the normalized amplitude of the interaction, and with $S=1/\prod_i(m_i!)$, where $m_i\in\mathbb{N}$ with $\sum m_i=n$ are the multiplicities of the output.

Observe that we have used here the same $\mathcal{M}$ symbol as in Principle 8.11. Obviously, this is something heavy, and deep, and the whole point with everything lies here. But in the lack of advanced physics tools in order to explain this, which as already mentioned on the occasion of Principle 8.11, come from quantum field theory, we will have to leave it like this, with of course the comment that "all this is verified by experiments".

Excuse me, but cat is here, meowing something. In English translation, what he says is not very funny, "with your lack of understanding of physics, that you practice every day, and share with your students too, you will end up destroying the whole planet". Well, dear cat, what can I answer. Physics is not everything in life, let's not forget about chemistry, and then especially biology, big and slow animals like dinosaurs, or nowadays humans, usually rule, and there is nothing much we can do about this.

Moving ahead now, with Principle 8.15 agreed upon, we can integrate as before over time, and we have the following analogue of Theorem 8.12, or Golden Rule 4:

Theorem 8.16 (Golden Rule 4'). *In the context of scattering, $*_a+*_b\to *_1+\ldots+*_n$, the cross section is given, in standard $\tilde{p}=(E/c,p)$ notation, by*

$$\begin{aligned}\sigma &= \frac{Sh^2}{4\sqrt{<\tilde{p}_a,\tilde{p}_b>^2-(m_am_bc^2)^2}}\int|\mathcal{M}|^2(2\pi)^4\delta\left(\tilde{p}_a+\tilde{p}_b-\sum_{i=1}^n\tilde{p}_i\right)\\ &\quad\times\prod_{i=1}^n\frac{1}{2\sqrt{||p_i||^2+m_i^2c^2}}\cdot\frac{dp_i}{(2\pi)^3}\end{aligned}$$

with $\mathcal{M}=\mathcal{M}(p_a,p_b,p_1,\ldots,p_n)$ being the normalized amplitude, $S=1/\prod_i(m_i!)$ being the statistical factor, and with the convention $E_i/c=\sqrt{||p_i||^2+m_i^2c^2}$, both in $\mathcal{M}$ and δ.

Proof. This is indeed nearly identical to the proof of Theorem 8.12, with all the manipulations there applying unchanged, to the present situation. ☐

Getting now to illustrations, let us work out the analogue of Theorem 8.13, by looking at situations of type $*_a+*_b\to *_1+*_2$. The result here is as follows:

THEOREM 8.17. *For events of type* $*_a + *_b \to *_1 + *_2$*, the Golden Rule for scattering, in the center of mass frame, gives the differential cross section formula*

$$\frac{d\sigma}{d\Omega} = \frac{Sh^2c^2|\mathcal{M}|^2}{64\pi^2(E_a+E_b)^2} \cdot \frac{||p_1||}{||p_a||}$$

with $\mathcal{M}$ *being the amplitude,* $||p_a|| = ||p_b||$ *and* $||p_1|| = ||p_2||$ *being the magnitudes of the incoming and outgoing momenta, and* $S = 1$ *if* $*_1 \neq *_2$*, and* $S = 1/2$ *if* $*_1 = *_2$*.*

PROOF. In the case $n = 2$, the formula in Theorem 8.16 takes the following form, with the statistical factor S being the one in the statement:

$$\begin{aligned}
\sigma &= \frac{Sh^2}{4\sqrt{<\tilde{p}_a,\tilde{p}_b>^2 - (m_am_bc^2)^2}} \int |\mathcal{M}|^2 \frac{(2\pi)^4\delta\left(\tilde{p}_a+\tilde{p}_b-\tilde{p}_1-\tilde{p}_2\right)}{4\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_2||^2+m_2^2c^2}} \cdot \frac{dp_1dp_2}{(2\pi)^6} \\
&= \frac{Sh^2}{64\pi^2\sqrt{<\tilde{p}_a,\tilde{p}_b>^2 - (m_am_bc^2)^2}} \int |\mathcal{M}|^2 \frac{\delta\left(\tilde{p}_a+\tilde{p}_b-\tilde{p}_1-\tilde{p}_2\right)}{\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_2||^2+m_2^2c^2}} \cdot dp_1dp_2
\end{aligned}$$

Our claim now is that, assuming that we are in the center of mass frame, where $p_b = -p_a$, the square root in the normalization factor is given by the following formula:

$$\sqrt{<\tilde{p}_a,\tilde{p}_b>^2 - (m_am_bc^2)^2} \simeq \frac{E_a+E_b}{c}\,||p_a||$$

With the notation $P = ||p_a||^2 = ||p_b||^2$, coming from our assumption $p_b = -p_a$, this is the same as proving that the following quantity is negligible:

$$K = c^2\left(<\tilde{p}_a,\tilde{p}_b>^2 - (m_am_bc^2)^2\right) - (E_a+E_b)^2P$$

In order to prove our claim, the first observation is that, according to our conventions for four-vectors, and to $p_b = -p_a$, we have the following formula:

$$<\tilde{p}_a,\tilde{p}_b> = \frac{E_aE_b}{c^2} - <p_a,p_b> = \frac{E_aE_b}{c^2} + P$$

We deduce from this that the above quantity K is given by:

$$\begin{aligned}
K &= c^2\left(\left(\frac{E_aE_b}{c^2}+P\right)^2 - (m_am_bc^2)^2\right) - (E_a+E_b)^2P \\
&= \frac{E_a^2E_b^2}{c^2} + c^2P^2 + 2E_aE_bP - m_a^2m_b^2c^6 - E_a^2P - E_b^2P - 2E_aE_bP \\
&= \frac{E_a^2E_b^2}{c^2} + c^2P^2 - m_a^2m_b^2c^6 - (E_a^2+E_b^2)P
\end{aligned}$$

In terms of Lorentz factors, we obtain from this the following formula:

$$\begin{aligned}
K &= \frac{(\gamma_a m_a c^2)^2(\gamma_b m_b c^2)^2}{c^2}+c^2P^2-m_a^2m_b^2c^6-\left((\gamma_a m_a c^2)^2+(\gamma_b m_b c^2)^2\right)P\\
&= \gamma_a^2\gamma_b^2 m_a^2m_b^2c^6+c^2P^2-m_a^2m_b^2c^6-\gamma_a^2m_a^2c^4P-\gamma_b^2m_b^2c^4P\\
&= (\gamma_a^2\gamma_b^2-1)m_a^2m_b^2c^6+c^2P^2-(\gamma_a^2m_a^2+\gamma_b^2m_b^2)c^4P
\end{aligned}$$

Now observe that, for a particle with relativistic momentum $p=\gamma mv$, the corresponding Lorentz factor γ can be computed by using the following formula:

$$\begin{aligned}
\sqrt{1+\left(\frac{||p||}{mc}\right)^2} &= \sqrt{1+\frac{\gamma^2||v||^2}{c^2}}\\
&= \sqrt{1+\frac{1}{1-||v||^2/c^2}\cdot\frac{||v||^2}{c^2}}\\
&= \sqrt{\frac{c^2}{c^2-||v||^2}}\\
&= \frac{1}{\sqrt{1-||v||^2/c^2}}\\
&= \gamma
\end{aligned}$$

By using this formula for both our particles $*_a$ and $*_b$, for which $||p_a||^2=||p_b||^2=P$, we can finish our computation of the above quantity K, and we obtain, as desired:

$$\begin{aligned}
K &= (\gamma_a^2\gamma_b^2-1)m_a^2m_b^2c^6+c^2P^2-(\gamma_a^2m_a^2+\gamma_b^2m_b^2)c^4P\\
&= \left(\left(1+\frac{P}{m_a^2c^2}\right)\left(1+\frac{P}{m_b^2c^2}\right)-1\right)m_a^2m_b^2c^6\\
&\quad -\left(\left(1+\frac{P}{m_a^2c^2}\right)m_a^2+\left(1+\frac{P}{m_b^2c^2}\right)m_b^2\right)c^4P\\
&= (m_a^2c^4P+m_b^2c^4P+c^2P^2)-(m_a^2c^4P+m_b^2c^4P+2c^2P^2)\\
&= -c^2P^2\\
&\simeq 0
\end{aligned}$$

Thus, claim proved, and with this in hand, our previous formula of the cross section, from the beginning of the present proof, takes the following form:

$$\sigma=\frac{Sh^2c}{64\pi^2(E_a+E_b)||p_a||}\int|\mathcal{M}|^2\frac{\delta\left(\tilde{p}_a+\tilde{p}_b-\tilde{p}_1-\tilde{p}_2\right)}{\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_2||^2+m_2^2c^2}}\cdot dp_1dp_2$$

Let us look now at the Dirac function. This decomposes over components, as follows:

$$\delta\left(\tilde{p}_a+\tilde{p}_b-\tilde{p}_1-\tilde{p}_2\right)=\delta\left(\frac{E_a}{c}+\frac{E_b}{c}-\frac{E_1}{c}-\frac{E_2}{c}\right)\delta\left(p_1+p_2\right)$$

Now recall from Theorem 8.16 that the machinery there leads to:

$$\frac{E_1}{c}=\sqrt{||p_1||^2+m_1^2c^2}\quad,\quad\frac{E_2}{c}=\sqrt{||p_2||^2+m_2^2c^2}$$

Thus, the above Dirac mass is in fact given by the following formula:

$$\delta\left(\tilde{p}_a+\tilde{p}_b-\tilde{p}_1-\tilde{p}_2\right)=\delta\left(\frac{E_a+E_b}{c}-\sqrt{||p_1||^2+m_1^2c^2}-\sqrt{||p_2||^2+m_2^2c^2}\right)\delta(p_1+p_2)$$

Getting back now to the formula of the cross section, that becomes:

$$\begin{aligned}\sigma &= \frac{Sh^2c}{64\pi^2(E_a+E_b)||p_a||}\int|\mathcal{M}|^2\frac{\delta\left(\frac{E_a+E_b}{c}-\sqrt{||p_1||^2+m_1^2c^2}-\sqrt{||p_2||^2+m_2^2c^2}\right)}{\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_2||^2+m_2^2c^2}}\\ &\quad\times\delta(p_1+p_2)\,dp_1dp_2\end{aligned}$$

Since we must have $p_2=-p_1$, this expression further simplifies to:

$$\sigma=\frac{Sh^2c}{64\pi^2(E_a+E_b)||p_a||}\int|\mathcal{M}|^2\frac{\delta\left(\frac{E_a+E_b}{c}-\sqrt{||p_1||^2+m_1^2c^2}-\sqrt{||p_1||^2+m_2^2c^2}\right)}{\sqrt{||p_1||^2+m_1^2c^2}\sqrt{||p_1||^2+m_2^2c^2}}\,dp_1$$

In spherical coordinates, this expression takes the following form:

$$\begin{aligned}\sigma &= \frac{Sh^2c}{64\pi^2(E_a+E_b)||p_a||}\int|\mathcal{M}|^2\frac{\delta\left(\frac{E_a+E_b}{c}-\sqrt{r^2+m_1^2c^2}-\sqrt{r^2+m_2^2c^2}\right)}{\sqrt{r^2+m_1^2c^2}\sqrt{r^2+m_2^2c^2}}\\ &\quad\times r^2\sin s\,drdsdt\end{aligned}$$

The problem now is that, with respect to what we did before for decay, something changes here in the physics, because the quantity $|\mathcal{M}|^2$ depends this time on the direction of p_1, and so we cannot do the angular integrations, trivially, as before. More on this later, when talking in detail about the amplitude $\mathcal{M}$, and in the meantime, let us finish our computation, by using some other methods and tricks. The idea will be very simple. Consider the solid angle Ω on the sphere, whose differential is given by:

$$d\Omega=\sin s\,dsdt$$

Now observe that we have the following formula, with the quantity on the left being something that we certainly want to compute, namely the differential cross section:

$$\frac{d\sigma}{d\Omega}=\frac{Sh^2c}{64\pi^2(E_a+E_b)||p_a||}\int_0^\infty|\mathcal{M}|^2\frac{\delta\left(\frac{E_a+E_b}{c}-\sqrt{r^2+m_1^2c^2}-\sqrt{r^2+m_2^2c^2}\right)}{\sqrt{r^2+m_1^2c^2}\sqrt{r^2+m_2^2c^2}}r^2dr$$

The point now is that the integral on the right is something familiar to us, and can be computed as before, in the decay case, in the proof of Theorem 8.13. To be more precise, following the proof there, consider the following quantity:

$$u=\sqrt{r^2+m_1^2c^2}+\sqrt{r^2+m_2^2c^2}$$

As before by differentiating, we obtain the following formula:

$$\frac{du}{dr}=\frac{ur}{\sqrt{r^2+m_1^2c^2}\sqrt{r^2+m_2^2c^2}}$$

Thus, in terms of this new variable u, we have the following formula:

$$\begin{aligned}\frac{d\sigma}{d\Omega}&=\frac{Sh^2c}{64\pi^2(E_a+E_b)||p_a||}\int_{m_1c+m_2c}^\infty|\mathcal{M}|^2\delta\left(\frac{E_a+E_b}{c}-u\right)\frac{r}{u}\,du\\&=\frac{Sh^2c^2r}{64\pi^2(E_a+E_b)^2||p_a||}\,|\mathcal{M}|^2\end{aligned}$$

Here in the last formula r stands for the value of the variable r evaluated at the place where the Dirac mass takes the value 1, that we can compute as follows:

$$\begin{aligned}&u=\frac{E_a+E_b}{c}\\\iff\quad&\sqrt{r^2+m_1^2c^2}+\sqrt{r^2+m_2^2c^2}=\frac{E_a+E_b}{c}\\\iff\quad&r^2+m_1^2c^2=r^2+m_2^2c^2+\left(\frac{E_a+E_b}{c}\right)^2-\frac{2(E_a+E_b)}{c}\sqrt{r^2+m_2^2c^2}\\\iff\quad&\frac{2(E_a+E_b)}{c}\sqrt{r^2+m_2^2c^2}=\left(\frac{E_a+E_b}{c}\right)^2+(m_2^2-m_1^2)c^2\\\iff\quad&2\sqrt{r^2+m_2^2c^2}=\frac{E_a+E_b}{c}+\frac{m_2^2-m_1^2}{E_a+E_b}\,c^3\\\iff\quad&4r^2+4m_2^2c^2=\left(\frac{E_a+E_b}{c}\right)^2+\left(\frac{m_2^2-m_1^2}{E_a+E_b}\right)^2c^6+2(m_2^2-m_1^2)c^2\\\iff\quad&4r^2=\left(\frac{E_a+E_b}{c}\right)^2+\left(\frac{m_2^2-m_1^2}{E_a+E_b}\right)^2c^6-2(m_1^2+m_2^2)c^2\\\iff\quad&r=\frac{1}{2}\sqrt{\left(\frac{E_a+E_b}{c}\right)^2+\left(\frac{m_2^2-m_1^2}{E_a+E_b}\right)^2c^6-2(m_1^2+m_2^2)c^2}\end{aligned}$$

Equivalently, with $r=||p||_1$, we are led to the formula in the statement. $\square$

8d. Amplitude, diagrams

With the above discussed, generalities of the Golden Rule in particle physics, for both decays and scattering, it remains to say what the amplitude $\mathcal{M}$ of the interactions is. With this being the missing ingredient, for what we have been doing, so far.

However, this is something quite tricky. To start with, we have been talking about decays $*_0 \to *_1 + \ldots + *_n$ and scattering $*_a + *_b \to *_1 + \ldots + *_n$, and the obvious good framework for discussing both situations is that of more complex events, as follows:

$$\circ_1 + \ldots + \circ_m \to *_1 + \ldots + *_n$$

Generally speaking, we can think at what happens here as being a quite general type of collision, with several input and output particles. So, let us draw a collision scheme, in the spirit of those what we drew some time ago, when doing classical mechanics:

$$\begin{tikzcd}
\circ_1 \arrow[dr] & \cdots & \circ_m \arrow[dl] \\
& \bigstar \arrow[dl] \arrow[dr] & \\
*_1 & \cdots & *_n
\end{tikzcd}$$

If you are already a bit familiar with Feynman diagrams, you will recognize here an order 0 such diagram, drawn a bit in an upside-down way. So, thanks for the remark, and in answer, we will be using in the remainder of this book the following convention:

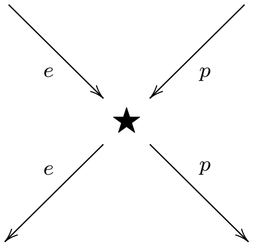
Convention 8.18. *For everything diagrams, we use the following conventions, which are commonplace in mathematics, and in modern theoretical physics too:*

1. *The vertical direction is for action, going from up to down.*
2. *The horizontal direction is for scalars, and other mathematics.*

We will be actually not using in this book scalars and other mathematics as in (2), which are rather advanced things, but it is useful I think to learn the basics in this way, so that you can understand later what modern mathematicians and physicists are doing, without being confused by notations and orientation. And as a joke here, to finish this discussion, all this is quite natural, because if we changed Dirac's notation for the scalar products, we should change Feynman's conventions for diagrams too, right.

Back to work now, we already have some knowledge about decays and scattering, in the particle physics context, but if there is something where we really do have some decent knowledge, that is the movement of the electron e around the proton p, in the context of the hydrogen atom, that we already spent most of the present book in studying.

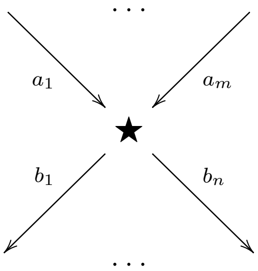
So, this will be our starting point, the mechanics of the hydrogen atom, with the electron e moving around the proton p. Obviously, what happens here is some sort of scattering, $e+p\to e+p$, so let us draw right away diagram for this situation:



You might say, end of the story here, this is the relevant diagram, and with that middle ★ symbol being actually a bit inappropriate, because nothing special happens anyway, electron e spinning smoothly around the proton p, with no fight or anything.

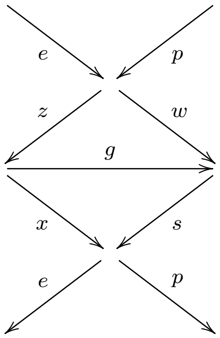
However, this is wrong. Remember all that corrections to the hydrogen atom, that we had troubles in understanding, and with what we understood being the tip of the iceberg, anyway? The whole point in advanced quantum mechanics lies there, in all that corrections, and quantum electrodynamics comes with the following bright idea:

IDEA 8.19. *Even for simple situations, like the hydrogen atom $e+p\to e+p$, the interactions should come in a hierarchic way, with the basic order 0 diagram*



being followed by order 1, order 2 and so on diagrams, and with the corresponding amplitude $\mathcal{M}$ being computed accordingly, as a power series in a certain variable α.

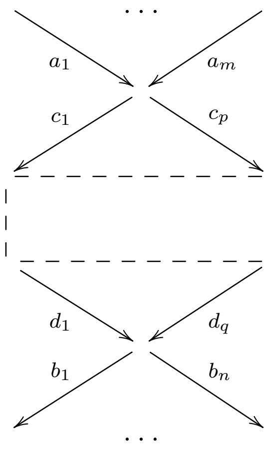
And isn't this bright. Everything illuminates now. So, we should blame all that mysterious corrections to the hydrogen atom on higher order interactions and diagrams, describing our $e+p\to e+p$ process, coming in complement to the order 0 diagram pictured above, with the typical higher order diagram being something as follows:



Here we have used some random letters, z,w,g,x,s, for the mysterious, very short lived particles that might appear at any time in our process $e+p\to e+p$, that we can hold responsible for the various corrections to the hydrogen atom, that we studied before. Of course, once our theory formally established, we will have to understand, via some further theory, and experiments too, what these beasts z,w,g,x,s really are.

In short, bright idea that we have, but enormous work still lying ahead, for making all this really work. Getting started now, inspired by the above, let us formulate:

Definition 8.20. *A Feynman diagram for a multiple scattering and decay event* $a_1+\ldots+a_m\to b_1+\ldots+b_m$ *is a diagram of type*



with c_i,d_i *being short-lived particles appearing in the event, and with the middle box being allowed to contain any such configuration of temporary particles too.*

Very nice all this, and getting now to work for good, so many things to be done. Let us start with the general recipe, since I am sure that you are very curious at this point about how all this works, and we will understand later what this really means:

Principle 8.21. *The amplitude $\mathcal{M}$ coming from a given Feynman diagram F can be computed as follows:*

1. *Label each vertex with the corresponding four-momentum vector.*
2. *Put factors $-ig$ at each vertex, g being the coupling constant.*
3. *Install Feynman propagators $\frac{i}{p_j^2-m_j^2c^2}$ on each internal line.*
4. *Install rescaled Dirac masses $(2\pi)^4\delta(\sum_i p_i)$ at each vertex.*
5. *Put integration factors $\frac{dp_j}{(2\pi)^4}$ on each internal line, and integrate.*
6. *Erase the global $(2\pi)^4\delta(\sum_i p_i)$ factor appearing after integrating.*
7. *Multiply the answer by i. That is your amplitude $\mathcal{M}$.*

Sounds exciting, doesn't it. Obviously, it will take us some time to understand, how all this works. To start with, passed the notations that we used in the above, which are a bit sloppy, but that we chose so in order not to overly complicate things, at least at this preliminary stage of things, what we are doing is quite simple, namely:

(1) The amplitude $\mathcal{M}$ appears by integrating over the state space, with conservation of energy and momentum being taken into account, at each internal vertex.

(2) What we are integrating are, basically, modulo some rescalings and other mathematical manipulations, the Feynman propagators, on each internal line.

(3) And with these mathematical manipulations including, crucially, the one at the end, namely erasing the final energy and momentum conservation term.

Which looks quite reasonable, physically speaking. However, a more careful look at Principle 8.21 reveals some sort of bug, coming from the coupling constant g appearing in (2) there. You would probably say, after all, what is this joke, we spent this whole chapter in waiting for the formula of the amplitude $\mathcal{M}$, and now here that formula comes, but by relegating everything to yet another beast, this coupling constant g.

Good point, and in answer right away, there is no cheating of any kind here, because we have the following very simple and precise rule, for quantum electrodynamics:

Rule 8.22. *In quantum electrodynamics, the coupling constant is*

$$g=\sqrt{4\pi\alpha}$$

with $\alpha\simeq 1/137$ being the fine structure constant.

In short, with this rule complementing Principle 8.21, that principle is meant to produce something numeric, as amplitude $\mathcal{M}$. We will see examples of this in a moment.

We have opted for stating Rule 8.22 independently of Principle 8.21, as a complement to it, because Principle 8.21 can work as well beyond quantum electrodynamics, for even finer theories, which require the use of other coupling constants g. More on this later.

As a basic illustration now for the above, chosen as simple as possible, without propagators, or integration work to be done, we have the following result:

THEOREM 8.23. *For a two-particle decay* $*_0 \to *_1 + *_2$ *the order* 0 *amplitude is* $\mathcal{M} = g$, *which gives via the Golden Rule an order* 0 *decay rate of*

$$\lambda = \frac{S||p||\alpha}{2hm_0^2c}$$

with $||p||$ *being the magnitude of either outgoing momentum,*

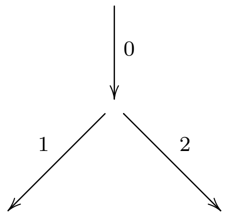
$$||p|| = \frac{c}{2m_0}\sqrt{m_0^4 + m_1^4 + m_2^4 - 2m_0^2m_1^2 - 2m_0^2m_2^2 - 2m_1^2m_2^2}$$

and the statistical factor being $S = 1$ *if* $*_1 \neq *_2$, *and* $S = 1/2$ *if* $*_1 = *_2$.

PROOF. We know from Theorem 8.13 that for two-particle decays, $*_0 \to *_1 + *_2$, the Golden Rule takes the following form, with S and $||p||$ being as in the statement:

$$\lambda = \frac{S||p||}{8\pi hm_0^2c}\,|\mathcal{M}|^2$$

In order to compute the amplitude $\mathcal{M}$, we use Principle 8.21. At order 0 we only have one Feynman diagram, which is the obvious one, namely:



Now let us apply Principle 8.21. We have a $-ig$ factor, no propagators, then a $(2\pi)^4\delta(p_0 - p_1 - p_2)$ factor which appears and dissapears, and so we get, right away:

$$\mathcal{M} = i(-ig) = g$$

Thus $|\mathcal{M}|^2 = g^2 = 4\pi\alpha$, which gives the formula of λ in the statement. □

The above example was of course extremely simple, without propagators, or integration work. However, do not worry, for decays at order 1, or for basic scatterings, $*_a + *_b \to *_1 + *_2$, such things will appear. We will discuss them later.

Summarizing, we have now some understanding of the fine structure of hydrogen.

8e. Exercises

Exercises:

Exercise 8.24.

Exercise 8.25.

Exercise 8.26.

Exercise 8.27.

Exercise 8.28.

Exercise 8.29.

Exercise 8.30.

Exercise 8.31.

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

9a. Particle physics

9b.

9c.

9d.

9e. Exercises

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

11a. Statistical mechanics

In order to advance, we need to talk about statistical mechanics. As a starting result here, dealing with the internal mechanism of the $PV = kT$ formula, we have:

THEOREM 11.1 (Maxwell). *The molecular speeds $v \in \mathbb{R}^3$ of a gas in thermal equilibrium are subject to the Maxwell-Boltzmann distribution formula*

$$P(v) = \left(\frac{m}{2\pi bT}\right)^{3/2} \exp\left(-\frac{m||v||^2}{2bT}\right)$$

with m being the mass of the molecules, and b being the Boltzmann constant.

PROOF. As usual with such things, this is rather something in between fact and theorem. Maxwell came upon it as a fact, or perhaps as a sort of pseudo-theorem, and a bit later Boltzmann came with a proof. In what follows we will discuss the original argument of Maxwell, then go towards Boltzmann's proof. Here is Maxwell's argument:

(1) We are looking for the precise probability distribution P of the molecular speeds $v = (v_1, v_2, v_3)$ which makes the mechanics of gases work. Intuition tells us that P has no corellations between the x, y, z directions of space, and so we must have:

$$P(v) = f(v_1)g(v_2)h(v_3)$$

Moreover, by rotational symmetry the functions f, g, h must coincide, and so:

$$P(v) = f(v_1)f(v_2)f(v_3)$$

(2) Further thinking, again invoking rotational symmetry, leads to the conclusion that $P(v)$ must depend only on the magnitude $||v||$ of the velocity $v \in \mathbb{R}^3$, and not on the direction. Thus, we must have as well a formula of the following type:

$$P(v) = \varphi(||v||^2)$$

(3) Now by comparing the requirements in (1) and (2), we are led via some math to the conclusion that φ must be an exponential, which amounts in saying that:

$$P(v) = \lambda \exp\left(-C||v||^2\right)$$

(4) Obviously we must have $C > 0$, for things to be bounded, and then by integrating we can obtain λ as function of C, and our formula becomes:

$$P(v) = \left(\frac{C}{\pi}\right)^{3/2} \exp\left(-C||v||^2\right)$$

(5) It remains to find the value of $C > 0$. But for this purpose, observe that, now that we have our distribution, be that still depending on $C > 0$, we can compute everything that we want to, just by integrating. In particular, we find that on average:

$$v_1^2 = v_2^2 = v_3^2 = \frac{1}{2C}$$

Thus the average magnitude of the molecular speed is given by:

$$||v|| = \frac{3}{2C}$$

It follows that the average kinetic energy of the molecules is:

$$K_0 = \frac{m||v||^2}{2} = \frac{3m}{4C}$$

(6) On the other hand, recall from basic thermodynamics that one of the many equivalent formulations of $PV = kT$, using $PV = 2K/3$, is as follows:

$$\frac{2K_0}{3} = bT$$

(7) Thus we obtain $m/(2C) = bT$, and so $C = m/(2bT)$, as desired. □

In order to further discuss all this, following Boltzmann, and then Gibbs and others, the idea will be that of connecting all the above to the key notion of entropy:

Theorem 11.2. *Given a gas, with states denoted $S = (P, V, T) \in \mathbb{R}^3$, subject to the equation of state $f(S) = 0$, define the entropy of a state by the formula*

$$\mathfrak{E}(S) = \int_{S_0}^{S} \frac{dQ}{T}$$

where S_0 is a chosen state, and the integral is over a reversible transformation from S_0 to S. Then $\mathfrak{E}$ is well-defined up to an additive scalar, and we have the inequality

$$\mathfrak{E}(S_1) \leq \mathfrak{E}(S_2)$$

for any transformation $S_1 \to S_2$, with equality when this transformation is reversible. That is, entropy increases, and there is nothing much that you can do about it.

Proof. This is something that we know well from basic thermodynamics, with the proof using the standard methods there, namely calculus, and Carnot machines. □

Following now Boltzmann, we have the following key result:

THEOREM 11.3 (Boltzmann). *The kinetic theory of gases, taking into account the collisions between molecules, leads to the formula for entropy*

$$\mathfrak{E}(S)=-b\int P(S)\log P(S)dS$$

with P being the probability on the state space, around our given state S.

PROOF. This is something quite complicated, and we refer to the literature. Observe that $\mathfrak{E}$ as computed above is given by an exact formula, not depending on an additive constant, the point here being that, in the context of Theorem 11.2, we can formally choose the state S_0 there to be the one, which is unique, at temperature $T=0$. □

Moving ahead now, as a second key result, also due to Boltzmann, we have:

THEOREM 11.4 (Boltzmann). *Given a gas with intial molecular speed distribution P, the collisions between molecules, leading to equilibrium, will work such as the quantity*

$$H=\int P(v)\log P(v)dv$$

decreases over time. The final distribution reached, over time, which is the one at equilibrium, is precisely the one which minimizes H, given by the formula

$$P(v)=\left(\frac{m}{2\pi bT}\right)^{3/2}\exp\left(-\frac{m||v||^2}{2bT}\right)$$

which is the Maxwell-Boltzmann distribution, that we knew before from Maxwell.

PROOF. Again, this is something at the same time a bit complicated, but of utter beauty, and for the proof, we again refer to any solid statistical mechanics book. □

Observe the obvious similarity between Theorem 11.3 and Theorem 11.4, and also the important theoretical consequences of the above, in view of the comments that we previously made in connection with the Maxwell-Boltzmann distribution.

In order to further clarify all the above, we will explain now a key result of Gibbs, called H theorem. This is something quite abstract and general, as follows:

THEOREM 11.5. *For a large system, wih states denoted S, the quantity*

$$H=\int P(S)\log P(S)dS$$

can only decrease over time, as to reach a minumum, where we have

$$P(S)\sim\exp\left(\sum_i\lambda_iE_i\right)$$

where E_i are the conserved quantities, and $\lambda_i\in\mathbb{R}$ are scalars.

Proof. This is something quite tricky, the idea being as follows:

(1) In order to prove this result, let us define quantities $A(S,T)$, depending on two states S,T, close to each other, such that the rate at which the system transitions $S\to T$ equals $A(S,T)dT$. With this convention, we have the following formula:

$$\frac{dP(S)dS}{dt}=\int\Big[P(T)A(T,S)dS-P(S)A(S,T)dS\Big]dT$$

By cancelling the differentials dS, this gives the following formula:

$$\frac{dP(S)}{dt}=\int\Big[P(T)A(T,S)-P(S)A(S,T)\Big]dT$$

(2) On the other hand, consider the quantity in the statement, namely:

$$H=\int P(S)\log P(S)dS$$

By using the formula found in (1), we obtain the following formula:

$$\frac{dH}{dt}=\int\int\Big[P(T)A(T,S)-P(S)A(S,T)\Big](\log P(S)+1)dSdT$$

Now by interchanging S,T in the second integral, this gives:

$$\frac{dH}{dt}=\int\int P(T)A(T,S)\log\frac{P(S)}{P(T)}\,dSdT$$

(3) We use now the following standard inequality, whose proof is elementary:

$$x,y\geq 0\implies y\log\frac{x}{y}\leq x-y$$

By applying this to the last formula found in (2), that formula gives:

$$\frac{dH}{dt}\leq\int\int(P(S)-P(T))A(T,S)dSdT$$

Now by interchanging again S,T, this inequality can be written as follows:

$$\frac{dH}{dt}\leq\int\int P(T)(A(S,T)-A(T,S))dSdT$$

(4) Normally this latter estimate allows us to conclude that H decreases, as claimed in the statement, because since the laws of physics must be invariant under time reversal, we should have $A(S,T)=A(T,S)$, and so our inequality simply reads:

$$\frac{dH}{dt}\leq 0$$

However, our argument has a flaw, in the context of advanced quantum mechanics, but as good news, the same advanced quantum mechanics tells us that we have:

$$\int(A(S,T)-A(T,S))dS=0$$

Thus, either way, we obtain $dH/dt\leq 0$, and so that H decreases, as claimed.

(5) In order to prove now the second assertion, regarding what happens when H is minimal, observe that for an infinitesimal change $\Delta P(S)$, we have:

$$\Delta H=\int\Delta P(S)(\log P(S)+1)dS$$

Now assuming that H is minimal, we must have $\Delta H=0$ for any allowed infinitesimal change $\Delta P(S)$, with this meaning that the change $\Delta P(S)$ must preserve the conserved quantities of the system, as for instance the energy $E(S)$, and also be such that:

$$\int P(S)dS=1$$

But this can only happen when $P(S)$ is of the following form, where E_i are the conserved quantities of the system, and $\lambda_i\in\mathbb R$ are certain scalars:

$$P(S)\sim\exp\left(\sum_i\lambda_iE_i\right)$$

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

In practice now, assuming that the energy E is the only conserved quantity, Theorem 11.5 leads to the following distribution, called "canonical ensemble":

$$P(S)=\exp\left[C-\frac{E(S)}{\theta}\right]$$

To be more precise, here θ is a certain constant, and C is another constant, uniquely determined by $\int P(S)dS=1$. But, in practice, we have $\theta=kT$, which gives:

$$P(S)=\exp\left[C-\frac{E(S)}{kT}\right]$$

More generally now, in case our system has other conserved quantities N_i, Theorem 11.5 leads to the following distribution, called "grand canonical ensemble":

$$P(S)=\exp\left[C-\frac{E(S)-\sum_i\lambda_iN_i(S)}{kT}\right]$$

We will be back to canonical ensembles later. For the moment, let us record the following result, coming as a complement to Theorem 11.5:

Theorem 11.6. *We have the formula*

$$\mathfrak{E}(S) = -kH$$

with $\mathfrak{E}$ being the thermodynamical entropy.

Proof. In the context of our equations above, adding some heat ΔQ to the system gives the following formula, for the corresponding change in the quantity H:

$$\begin{aligned}\Delta H &= \int \Delta P(S)(\log P(S)+1)dS\\ &= -\frac{1}{kT}\int \Delta P(S)E(S)dS\\ &= -\frac{\Delta Q}{kT}\end{aligned}$$

Now since the thermodynamical entropy satisfies by definition $d\mathfrak{E} = dQ/kT$, we conclude from the above equation that we have $\mathfrak{E}(S) = -kH$, as claimed. □

11b.

11c.

11d.

11e. Exercises

Exercises:

Exercise 11.7.

Exercise 11.8.

Exercise 11.9.

Exercise 11.10.

Exercise 11.11.

Exercise 11.12.

Exercise 11.13.

Exercise 11.14.

Bonus exercise.

CHAPTER 12

Exotic matter

12a. Exotic matter

12b.

12c.

12d.

12e. Exercises

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

13a. Periodic table

Let us investigate now the case of arbitrary atoms. We will need some general theory for the many-particle systems in quantum mechanics. Let us start with:

Definition 13.1. *The wave function of a system of electrons $e_1,\ldots,e_Z$, given by*

$$P_t\big(e_1\in V_1,\ldots,e_Z\in V_Z\big)=\int_{V_1\times\ldots\times V_Z}|\psi_t(x_1,\ldots,x_Z)|^2dx$$

is governed by the Schrödinger equation $ih\dot{\psi}=\widehat{H}\psi$, with Hamiltonian as follows,

$$\widehat{H}=-\frac{h}{2m}\sum_i\Delta_i+Ke^2\sum_{i<j}\frac{1}{||x_i-x_j||}+V(x_1,\ldots,x_Z)$$

with the middle sum standing for the Coulomb repulsions between them.

As before with the one-particle Schrödinger equation, there is a long story with all this, and for cutting short with the discussion here, this is what experiments lead to.

In general, and in fact at any $Z>1$, and so even at $Z=2$, the above Schrödinger equation is pretty much impossible to solve, due to the Coulomb repulsion term, which makes the mathematics extremely complicated. In fact, as an illustrating analogy here, managing that Coulomb repulsion term is more or less the same thing as solving the N-body problem in classical mechanics, for bodies with equal mass.

We will be interested here in the case of atoms, where V is the Coulomb attraction potential coming from a Ze charge. Here the problem to be solved is as follows:

Problem 13.2. *Consider an atom of atomic number Z, meaning a fixed Ze charge, surrounded by electrons $e_1,\ldots,e_Z$. The problem is to solve the Schrödinger equation*

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

with Hamiltonian as follows,

$$\widehat{H}=\sum_i\left(-\frac{h}{2m}\Delta_i-\frac{KZe^2}{||x_i||}\right)+Ke^2\sum_{i<j}\frac{1}{||x_i-x_j||}$$

or at least to understand how $e_1,\ldots,e_Z$ manage to live together, in a stable way.

A first idea would be of course that of ignoring the right term, Coulomb repulsion. In the simplest case, that of the helium atom, the situation is as follows:

FACT 13.3. *For the helium atom, $Z = 2$, ignoring the Coulomb repulsion between electrons leads, via separation of variables, to product wave functions*

$$\phi(x_1, x_2) = \phi'_{n_1 l_1 m_1}(x_1)\phi'_{n_2 l_2 m_2}(x_2)$$

with the prime signs standing for the doubling $e \to 2e$ of the central charge, with energies:

$$E_{n_1 n_2} = 4(E_{n_1} + E_{n_2})$$

This model predicts a ground state energy for helium given by

$$E_0 = 8 \times (-13.6) = -109 \text{ eV}$$

which is considerably smaller than the observed $E_0 = -79$ eV.

As a partial conclusion to what we have so far, things not going on every well, and in order to advance, we will probably need to invest a lot of time in learning how to solve complicated Schrödinger equations, and why not buying a super-computer too.

Moving ahead, let us focus on a more modest question, that at the end of Problem 13.2, namely understanding how the electrons $e_1, \ldots, e_Z$ manage to live together. Here our method of ignoring the Coulomb repulsion between electrons is not that bad, and for helium for instance, we are led in this way to some interesting conclusions. For instance the excited states of helium must appear as products as follows:

$$\phi_{100}(x_1)\phi_{nlm}(x_2) \quad , \quad \phi_{nlm}(x_1)\phi_{100}(x_2)$$

And this is actually quite close to reality, if we add the particle spin to our discussion. So, speaking now spin, this is indeed something that we ignored so far in the above. And spin is in fact a key component to our problem, because we have:

FACT 13.4 (Pauli exclusion principle). *Two electrons cannot occupy the same quantum numbers n, l, m, with same spin $s = \pm 1/2$.*

So, this is the famous Pauli exclusion principle, giving the golden key to the understanding of $Z \geq 2$ atoms. There are of course many things that can be said about it. A sample quantum mechanics book will probably tell you first something about bosons and fermions, coming with exactly 0 evidence, then some more things about electrons, of type “they are the same, but not really, and everything is entangled, but is it really entangled”, and finally formulate the Pauli exclusion principle, as a theorem.

We will not get into this here, and take the Pauli exclusion principle as it is, a physics fact. However, talking philosophy, personally I always think at it as coming from the “human nature of electrons”. To be more precise, when I'm at my office typing the present book, I feel like occupying some precise quantum numbers, with precise spin.

And if 30 colleagues, all typing physics books too, manage to come by surprise to my office, and squeeze there like sardines, I will surely find a way of getting rid of them, and disposing of their bodies. So now that a poor human like me can do this, why shouldn't a mighty electron be able to do the same. This is the Pauli exclusion principle.

We have now all the ingredients for discussing the known atoms, or chemical elements, $Z = 1, \ldots, 118$. These can be arranged in a table, called periodic table, as follows:

	1	2		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	H 1																		He 2
2	Li 3	Be 4												B 5	C 6	N 7	O 8	F 9	Ne 10
3	Na 11	Mg 12												Al 13	Si 14	P 15	S 16	Cl 17	Ar 18
4	K 19	Ca 20		Sc 21	Ti 22	V 23	Cr 24	Mn 25	Fe 26	Co 27	Ni 28	Cu 29	Zn 30	Ga 31	Ge 32	As 33	Se 34	Br 35	Kr 36
5	Rb 37	Sr 38		Y 39	Zr 40	Nb 41	Mo 42	Tc 43	Ru 44	Rh 45	Pd 46	Ag 47	Cd 48	In 49	Sn 50	Sb 51	Te 52	I 53	Xe 54
6	Cs 55	Ba 56	l	Lu 71	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78	Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Po 84	At 85	Rn 86
7	Fr 87	Ra 88	a	Lr 103	Rf 104	Db 105	Sg 106	Bh 107	Hs 108	Mt 109	Ds 110	Rg 111	Cn 112	Nh 113	Fl 114	Mc 115	Lv 116	Ts 117	Og 118
			l:	La 57	Ce 58	Pr 59	Nd 60	Pm 61	Sm 62	Eu 63	Gd 64	Tb 65	Dy 66	Ho 67	Er 68	Tm 69	Yb 70		
			a:	Ac 89	Th 90	Pa 91	U 92	Np 93	Pu 94	Am 95	Cm 96	Bk 97	Cf 98	Es 99	Fm 100	Md 101	No 102		

Here the horizontal parameter $1, \ldots, 18$ is called the group, and the vertical parameter $1, \ldots, 7$ is called the period. The two bottom rows consist of lanthanum $_{57}\mathrm{La}$ and its followers, called lanthanides, and of actinium $_{89}\mathrm{Ac}$ and its followers, called actinides.

Thus, the periodic table, when correctly drawn, but no one does that because of obvious typographical reasons, is in fact a 7×32 table. Note here that, according to our 7×18 convention, which is the standard one, lanthanides and actinides don't have a group number $1, \ldots, 18$. Their group is by definition "lanthanides" and "actinides".

We will comment in a moment on all this, but before anything:

ADVICE 13.5. *Learn their names.*

This is a serious advice, the periodic table being the main theorem of mathematics, physics, chemistry, biology and engineering combined. So if there's one theorem to be learned, full statement, that is the one. In case you're out of memory, just erase from your brain everything that you learned so far from this book of mine, and learn instead that 118 elements. Please do it for me, this being my final wish, from the death bed.

Actually, in order to get started here, here are the names up to krypton ${}_{36}$Kr, which are absolutely needed for everything, and must be all learned, to start with:

Definition 13.6. *The elements up to krypton* ${}_{36}$Kr *are as follows:*

1. *Hydrogen* ${}_{1}$H, *helium* ${}_{2}$He.
2. *Lithium* ${}_{3}$Li, *beryllium* ${}_{4}$Be, *boron* ${}_{5}$B, *carbon* ${}_{6}$C, *nitrogen* ${}_{7}$N, *oxygen* ${}_{8}$O, *fluorine* ${}_{9}$F, *neon* ${}_{10}$Ne.
3. *Sodium* ${}_{11}$Na, *magnesium* ${}_{12}$Mg, *aluminium* ${}_{13}$Al, *silicon* ${}_{14}$Si, *phosphorus* ${}_{15}$P, *sulfur* ${}_{16}$S, *chlorine* ${}_{17}$Cl, *argon* ${}_{18}$Ar.
4. *Potassium* ${}_{19}$K, *calcium* ${}_{20}$Ca, *scandium* ${}_{21}$Sc, *titanium* ${}_{22}$Ti, *vanadium* ${}_{23}$V, *chromium* ${}_{24}$Cr, *manganese* ${}_{25}$Mn, *iron* ${}_{26}$Fe, *cobalt* ${}_{27}$Co.
5. *Nickel* ${}_{28}$Ni, *copper* ${}_{29}$Cu, *zinc* ${}_{30}$Zn, *gallium* ${}_{31}$Ga, *germanium* ${}_{32}$Ge, *arsenic* ${}_{33}$As, *selenium* ${}_{34}$Se, *bromine* ${}_{35}$Br, *krypton* ${}_{36}$Kr.

Observe that all names fit with the abbreviations, expect for sodium ${}_{11}$Na, coming from the Latin natrium, potassium ${}_{19}$K, coming from the Latin kalium, iron ${}_{26}$Fe coming from the Latin ferrum, and also copper ${}_{29}$Cu, coming from the Latin cuprum.

In what regards the elements heavier than krypton ${}_{36}$Kr, it is heartbreaking to sort them out, but as a useful complement to the above list, we have:

Fact 13.7. *Remarkable elements heavier than krypton* ${}_{36}$Kr *include:*

1. *Noble gases: xenon* ${}_{54}$Xe, *radon* ${}_{86}$Rn.
2. *Noble metals: silver* ${}_{47}$Ag, *iridium* ${}_{77}$Ir, *platinum* ${}_{78}$Pt, *gold* ${}_{47}$Au.
3. *Heavy metals: mercury* ${}_{80}$Hg, *lead* ${}_{82}$Pb.
4. *Radioactive: polonium* ${}_{84}$Po, *radium* ${}_{88}$Ra, *uranium* ${}_{92}$U, *plutonium* ${}_{94}$Pu.
5. *Miscellaneous: rubidium* ${}_{37}$Rb, *strontium* ${}_{38}$Sr, *molybdenum* ${}_{42}$Mo, *technetium* ${}_{43}$Tc, *cadmium* ${}_{48}$Cd, *tin* ${}_{50}$Sn, *iodine* ${}_{53}$I, *caesium* ${}_{55}$Cs, *tungsten* ${}_{74}$Tu, *bismuth* ${}_{83}$Bi, *francium* ${}_{87}$Fr, *americium* ${}_{95}$Am.

Here the abbreviations not fitting with English names come from the Latin or sometimes Greek argentum ${}_{47}$Ag, aurum ${}_{47}$Au, hydrargyrum ${}_{80}$Hg, plumbum ${}_{82}$Pb and stannum ${}_{50}$Sn. The noble gases in (1) normally include oganesson ${}_{118}$Og as well. The noble metals in (2) are something subjective. There are of course plenty of other heavy metals (3), or radioactive elements (4). As for the list in (5), this is something subjective, basically a mixture of well-known metals used in engineering, and some well-known bad guys in the context of nuclear fallout. Technetium ${}_{43}$Tc is a bizarre element, human-made.

But let us not forget about quantum mechanics, and what we wanted to do, namely discuss electron structure. As a first observation, we have:

FACT 13.8. *Any* $Z = 1, \ldots, 118$ *corresponds to a unique element, having* Z *protons in the core, and* Z *electrons around it. This element might come with isotopes, depending on the number of neutrons in the core, can be in ground state or excited states, can get ionized, and so on, but all there versions are "family", and the element is unique.*

This fact is something which might look very natural, with no need for explanation for it, but after some thinking, is this really that natural. And the answer here is that no, if you don't know quantum mechanics, and yes, if you know some, as we do:

(1) For the purposes of our question, we can assume that we are in the context of Problem 13.2, and with the Coulomb repulsions between electrons ignored.

(2) But then, we are a bit in the same situation as in Fact 13.3, and the analysis there, based on hydrogen theory modified via $e \to Ze$, carries over.

(3) And so, the Z electrons will arrange on various energy levels, subject to Pauli exclusion, as to occupy a state of lowest possible energy, so the solution is unique.

In fact, we can now understand the electron structure of the various elements, and also how the periodic table is exactly made, the conclusions here being as follows:

FACT 13.9. *For the element having atomic number* Z*, the electrons will occupy successively the various positions with quantum numbers* $n, l, m \in \mathbb{N}$ *and spin* $s = \pm 1/2$*, such as the total binding energy to be minimal. In practice, the period* $1, \ldots, 7$ *corresponds to the highest* n *occupied, and the group* $1, \ldots, 18$ *comes from* l, m, s*.*

This is of course something very basic, and there is a detailed analysis to be done afterwards, for $Z = 1, \ldots, 118$. For the elements up to krypton ${}_{36}$Kr, the list of electron configurations can be found for instance in Feynman [**35**] or Griffiths [**43**].

Time now for more advanced mathematics, for the many-electron atoms, and by talking about isotopes and ionization too, and this even for the hydrogen atom. All this knowledge will be very useful, among others for going towards molecules, afterwards.

In other words, this means that we will be interested in what happens to a system of Z electrons $e_1, \ldots, e_Z$, surrounding a central positive charge $Z'e$. For a usual atom, which is globally electrically neutral, we have $Z = Z'$, but for isotopes and ions we can have $Z < Z'$ or $Z > Z'$. Thus, we will assume that the numbers Z, Z' are unrelated.

In practice, this corresponds to the following version of Problem 13.2:

Problem 13.10. *Consider a system of atomic number Z', meaning a fixed $Z'e$ charge, surrounded by electrons $e_1,\ldots,e_Z$. The problem is to solve the Schrödinger equation*

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

with Hamiltonian as follows,

$$\widehat{H}=\sum_i\left(-\frac{h}{2m}\,\Delta_i-\frac{KZ'e^2}{||x_i||}\right)+Ke^2\sum_{i<j}\frac{1}{||x_i-x_j||}$$

or at least to understand how $e_1,\ldots,e_Z$ manage to live together, in a stable way.

As explained in the discussion following Problem 13.2, a first idea is to simply ignore the Coulomb repulsion term on the right. Indeed, this simplifies a lot the mathematics, and by separation of variables we are led to a product of wave functions, with the numerics being worked out, in the simplest case of the helium atom, in Fact 13.3.

So, let us see how this works in general, in the framework of Problem 13.10. As before with helium, in view of the fact that the interactions between the electrons are ignored, this amounts in decomposing the Hamiltonian into Z components, as follows:

$$\widehat{H}=\widehat{H}_1+\ldots+\widehat{H}_Z$$

By separation of variables, we are led to products of wave functions as follows, called Hartree products, with the prime signs standing for the modification of the central charge, $e\to Z'e$, from the case of hydrogen, to the case of the system under investigation:

$$\phi(x_1,\ldots,x_Z)=\phi'(x_1)\ldots\phi'(x_Z)$$

Here we have opted, for simplifying notations, to not include the quantum numbers, as in Fact 13.3, at least in the present, preliminary stage of our study.

With this done, we are quite far from something reliable, because as explained in Fact 13.3, such an approximation gives quite average results, with respect to the observed values, even in the simplest case of helium. So, the following question appears:

Question 13.11. *How to further improve the Hartree products, without however getting into Problem 13.10 as stated, which is something of extreme difficulty?*

In order to solve this question, let us get back to helium, $Z=Z'=2$. If we denote for simplifying by $1,2$ the first two lowest energy orbits, we have two possible Hartree functions for the simplest excited state of our helium atom, namely:

$$\phi_{12}(x_1,x_2)=\phi_1'(x_1)\phi_2'(x_2)\quad,\quad\phi_{21}(x_1,x_2)=\phi_2'(x_1)\phi_1'(x_2)$$

A natural idea, in order to have some symmetry going on, for our solution, is that of considering a suitable linear combination of these solutions. But since the overall electron density function $|\phi|^2$ must be invariant under electron exchange, we are led, up to a normalization of the wave function, to linear combinations as follows:

$$\phi = \pm\phi_{12} \pm \phi_{21}$$

Moreover, by taking now into account spin, the Pauli exclusion principle tells us that the correct symmetry property of ϕ is actually antisymmetry. Thus, up to a global ± sign, and again up to a normalization of the wave function, the solution must be:

$$\begin{aligned}\phi &= \phi_{12} - \phi_{21} \\ &= \phi_1'(x_1)\phi_2'(x_2) - \phi_2'(x_1)\phi_1'(x_2) \\ &= \begin{vmatrix}\phi_1'(x_1) & \phi_2'(x_1) \\ \phi_1'(x_2) & \phi_2'(x_2)\end{vmatrix}\end{aligned}$$

Getting now to the normalization factor, a simple computation shows that this factor is $1/\sqrt{2}$. Thus, as a conclusion, our symmetrization method leads to:

$$\phi = \frac{1}{\sqrt{2}}\begin{vmatrix}\phi_1'(x_1) & \phi_2'(x_1) \\ \phi_1'(x_2) & \phi_2'(x_2)\end{vmatrix}$$

More generally now, the above method applies in the same way to a system of Z electrons, and we are led to the following preliminary answer to Question 13.11:

Answer 13.12. *The correct linear combinations of Hartree products, having the correct antisymmetrization property for the electrons, are the quantities*

$$\phi = \frac{1}{\sqrt{Z!}}\begin{vmatrix}\phi_1'(x_1) & \ldots & \phi_Z'(x_1) \\ \vdots & & \vdots \\ \phi_1'(x_Z) & \ldots & \phi_Z'(x_Z)\end{vmatrix}$$

with the subscripts standing for the hydrogen-like energy levels, and the primes standing for the central charge modification $e \to Z'e$*, called Slater determinants.*

To be more precise here, the fact that we must indeed consider a determinant is standard, by reasoning as above, and with this actually corresponding to a well-known theorem in mathematics, stating that the determinant is the unique antisymmetric multilinear form $\det : \mathbb{R}^N \to \mathbb{R}$, normalized as to produce 1 for the standard basis of $\mathbb{R}^N$. As for the computation of the normalization factor, this is again standard, as above.

Moving ahead now, the electron spin was certainly taken into account when formulating the above answer, due to the Pauli exclusion principle which was used. However, when fully taking spin into account, we are led to the following refinement of the above formula, valid this time for a system of $Z = 2N$ electrons, with the subscripts ignoring spin, and with the bars, and lack of bars, standing for spin up and down:

$$\phi=\frac{1}{\sqrt{Z!}}\begin{vmatrix}\phi_1'(x_1)&\bar{\phi}_1'(x_1)&\ldots&\ldots&\phi_N'(x_1)&\bar{\phi}_N'(x_1)\\ \phi_1'(x_2)&\bar{\phi}_1'(x_2)&\ldots&\ldots&\phi_N'(x_2)&\bar{\phi}_N'(x_2)\\ \vdots&\vdots&&&\vdots&\vdots\\ \vdots&\vdots&&&\vdots&\vdots\\ \phi_1'(x_{Z-1})&\bar{\phi}_1'(x_{Z-1})&\ldots&\ldots&\phi_N'(x_{Z-1})&\bar{\phi}_N'(x_{Z-1})\\ \phi_1'(x_Z)&\bar{\phi}_1'(x_Z)&\ldots&\ldots&\phi_N'(x_Z)&\bar{\phi}_N'(x_Z)\end{vmatrix}$$

However, we will not get into full details here, because an even better approximation method, called Hartree-Fock approximation, beating the above, is still to come.

As an example, however, for all this, let us discuss the case of the beryllium atom $_4$Be. Here the Slater determinant, taking into account spin, as above, is as follows:

$$\phi=\frac{1}{\sqrt{24}}\begin{vmatrix}\phi_1'(x_1)&\bar{\phi}_1'(x_1)&\phi_2'(x_1)&\bar{\phi}_2'(x_1)\\ \phi_1'(x_2)&\bar{\phi}_1'(x_2)&\phi_2'(x_2)&\bar{\phi}_2'(x_2)\\ \phi_1'(x_3)&\bar{\phi}_1'(x_3)&\phi_2'(x_3)&\bar{\phi}_2'(x_3)\\ \phi_1'(x_4)&\bar{\phi}_1'(x_4)&\phi_2'(x_4)&\bar{\phi}_2'(x_4)\end{vmatrix}$$

Getting now to the excited states of the same beryllium $_4$Be, that we would like to understand now, we need to add here a third orbital, that we will label 3.

And then, following a discussion about spin, which must be subject to certain rules, we are led to the conclusion that the correct linear combination of Hartree products is a difference of two Slater determinants, as follows:

$$\begin{aligned}\phi\ &=\ \frac{1}{\sqrt{24}}\begin{vmatrix}\phi_1'(x_1)&\bar{\phi}_1'(x_1)&\phi_2'(x_1)&\bar{\phi}_3'(x_1)\\ \phi_1'(x_2)&\bar{\phi}_1'(x_2)&\phi_2'(x_2)&\bar{\phi}_3'(x_2)\\ \phi_1'(x_3)&\bar{\phi}_1'(x_3)&\phi_2'(x_3)&\bar{\phi}_3'(x_3)\\ \phi_1'(x_4)&\bar{\phi}_1'(x_4)&\phi_2'(x_4)&\bar{\phi}_3'(x_4)\end{vmatrix}\\ &-\ \frac{1}{\sqrt{24}}\begin{vmatrix}\phi_1'(x_1)&\bar{\phi}_1'(x_1)&\bar{\phi}_2'(x_1)&\phi_3'(x_1)\\ \phi_1'(x_2)&\bar{\phi}_1'(x_2)&\bar{\phi}_2'(x_2)&\phi_3'(x_2)\\ \phi_1'(x_3)&\bar{\phi}_1'(x_3)&\bar{\phi}_2'(x_3)&\phi_3'(x_3)\\ \phi_1'(x_4)&\bar{\phi}_1'(x_4)&\bar{\phi}_2'(x_4)&\phi_3'(x_4)\end{vmatrix}\end{aligned}$$

As already mentioned, we will not get into full details here, because an even better approximation method, called Hartree-Fock approximation, is still to come.

Central field approximation.

Adding spin, Hartree-Fock approximation.

Many things can be said here, with this being a quite powerful method.

Good news, with what we know from the above about atoms, and in particular with the knowledge of the Hartree-Fock method, we can now talk about molecules.

Indeed, the idea here is that a molecule can be investigated a bit like a multi-electron atom, by replacing the central charge with a system of positive charges.

In practice, by further building on the Hartree-Fock method for the atoms, we are led in this way to the Born-Oppenheimer method for the molecules.

Many things can be said here, notably with a number of more advanced models for the simplest molecule of them all, which is the hydrogen molecule.

With what we know about molecules, we can now start doing some chemistry.

There is an enormous quantity of things that can be said here, all relying on quite delicate mathematics, the simplest of which being as follows:

Theorem 13.13. *The group* 18 *elements, namely*

(1) *Helium* ${}_2\mathrm{He}$*,*
(2) *Neon* ${}_{10}\mathrm{Ne}$*,*
(3) *Argon* ${}_{18}\mathrm{Ar}$*,*
(4) *Krypton* ${}_{36}\mathrm{Kr}$*,*
(5) *Xenon* ${}_{54}\mathrm{Xe}$*,*
(6) *Radon* ${}_{86}\mathrm{Rn}$*,*

called noble gases, are allergic to chemistry.

Proof. This follows from the electron structure, because the group 18 elements are precisely those with all possible electron positions fully occupied, up to a certain $n \in \mathbb{N}$, which makes them very unfriendly to any chemistry proposition from the outside. By the way, oganesson ${}_{118}\mathrm{Og}$ is normally part of this group too, but since this element has only been created and observed for a tiny fraction of a second, who really knows, and by the standard scientific etiquette, in the lack of experiments, no comment about it. $\square$

More about noble gases.

The particular case of helium ${}_2\mathrm{He}$, which is a remarkable element.

Discussion about radon ${}_{86}\mathrm{Rn}$ too, going beyond basic physics and chemistry.

13b.

13c.

13d.

13e. Exercises

Exercises:

Exercise 13.14.

Exercise 13.15.

Exercise 13.16.

Exercise 13.17.

Exercise 13.18.

Exercise 13.19.

Exercise 13.20.

Exercise 13.21.

Bonus exercise.

CHAPTER 14

Ions, isotopes

14a. Ions, isotopes

14b.

14c.

14d.

14e. Exercises

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

15a. Small molecules

15b.

15c.

15d.

15e. Exercises

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

16a. Big molecules

16b.

16c.

16d.

16e. Exercises

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