

Methods of electrodynamics

Teo Banica

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CERGY-PONTOISE, F-95000
CERGY-PONTOISE, FRANCE. teo.banica@gmail.com

2010 *Mathematics Subject Classification.* 81R50

Key words and phrases. Electrodynamics, Free electrodynamics

ABSTRACT. This is an introduction to classical and quantum electrodynamics, starting from the very basics, and assuming calculus known, and then to some modern, speculative versions of the theory, namely the twisted version, and the free version.

Preface

Electrodynamics is to modern physics what classical mechanics was to classical physics. Everything comes from there. A careful study of the Maxwell equations shows that these are invariant under a tricky transformation of the space-time, called Lorentz transformation, and by further thinking along these lines, you can come upon relativity. Also, the 2-body problem in electrodynamics, namely how an electron spins around a proton, leads to quantum mechanics, and eventually to a better version of electrodynamics, called quantum electrodynamics. Finally, light is an electromagnetic phenomenon, and so is heat, and all this is essential for the understanding of modern thermodynamics.

Basic electrodynamics can be learned from many places, and my personal absolute favorite is Griffiths' book [44]. What a book. In fact, looking at my own collection of books, a few hundred items, all sorts of mathematics and physics texts, if there's a book there which is really good, definitely worth the money, that is Griffiths'.

Basic electrodynamics, as explained in [44], is of course not everything. Remember what happened to classical mechanics, which proved to be not exactly correct, and was eventually updated into general relativity? Something similar happened to electrodynamics, which was eventually updated into quantum electrodynamics, a theory invented by Feynman and others, which incorporates quantum mechanics, and correctly solves the 2-body problem. Again, there are many good books here, starting with Feynman's own book [40], coming as a continuation of his lovely physics course [37], [38], [39]. You can also learn this from Griffiths [45], [46], or from many other places.

Of course things are not over either with quantum electrodynamics, first because a tantalizing open question, regarding the precise nature of the main number appearing there, in all formulae, which is the fine structure constant $\alpha \simeq \frac{1}{137}$. This is a dimensionless constant, appearing as a product of various household constants in physics, and finding an exact mathematical formula for it is one of the main problems in modern science.

As a second open question, one problem is that of zooming down, far beyond the level of electrons, protons and neutrons, which produce quantum electrodynamics, as to get to smaller particles, which can only produce a better version of electrodynamics. This is something that has been heavily tried in the last 50-60 years, with all sorts of beasts

found there, such as quarks, strings, quantum fields and more. However, at least so far, at the level of quantum electrodynamics, the Feynman version of the theory still remains strong, and is rather unimpressed by all these modern beasts. But, no reasons to be pessimistic here, the correct beasts, beating Feynman, will certainly emerge one day. For some reading on all this, you have Weinberg, [89], [90], [91], [92] and more.

As a conclusion to all this, electrodynamics, in its various incarnations, is a rather heavy theory, with hundreds of people working on this, now as I type, or as you read. Also, at the level of books and presentation, basically everyone knowing a bit about it, starting with those who had developed the theory, big names in physics, having a deep knowledge of the subject, has tried, and the stakes are considerably high.

The present book is, you guessed it, about electrodynamics. Normally you need a good reason, or rather excuse, for writing a new book here, and my own excuse is that I would like to present, besides the basics, some modern speculations on the subject:

Parts I-II are an introduction to classical and quantum electrodynamics, starting from the very basics, and assuming calculus known. These were certainly a pleasure to write, of rather guilty type, and in the hope that I got this right, not quite sure about this. For more, or better, go of course with Feynman, Griffiths, Weinberg and the others.

Parts III-IV are an introduction to various versions of the theory, which are rather mathematical, coming from recent work on beasts such as quantum groups, and quantum manifolds. We will first discuss a twisted version of quantum electrodynamics, and then a free version of it, formally related to quantum chromodynamics and quarks.

It is a pleasure to thank my mathematics and physics colleagues, so much to learn from them, over the time. Many thanks as well to those who have contributed to the quantum group and quantum manifold material at the end. Finally, many thanks go to my cats. Their fine knowledge of life, matter, and of $\alpha \simeq \frac{1}{137}$, of course not shared with omnivorous creatures like me, has always impressed and guided me.

Contents

Preface	3
Part I. Classical electrodynamics	9
Chapter 1. Coulomb, Gauss	11
1a. Point charges	11
1b. Field lines	15
1c. The flux	18
1d. Gauss law	21
1e. Exercises	22
Chapter 2. Poisson and Laplace	23
2a. Some calculus	23
2b. Gauss, again	26
2c. Poisson equation	27
2d. Faraday cage	30
2e. Exercises	30
Chapter 3. Maxwell equations	31
3a. Magnetism	31
3b. Electric currents	33
3c. Faraday, Maxwell	40
3d. Lorentz invariance	43
3e. Exercises	44
Chapter 4. Radiation and waves	45
4a. Wave equation	45
4b. Radiation, light	47
4c. Spectral lines	52
4d. Max Planck	61
4e. Exercises	66

Part II. Quantum electrodynamics	67
Chapter 5. Quantum mechanics	69
5a. Schrödinger equation	69
5b. Spherical coordinates	72
5c. Spherical harmonics	77
5d. Bohr energy	82
5e. Exercises	88
Chapter 6. Fine structure	89
6a. Fine structure	89
6b.	92
6c.	92
6d.	92
6e. Exercises	92
Chapter 7. Feynman diagrams	93
7a.	93
7b.	93
7c.	93
7d.	93
7e. Exercises	93
Chapter 8. Renormalization	95
8a.	95
8b.	95
8c.	95
8d.	95
8e. Exercises	95
Part III. Twisted electrodynamics	97
Chapter 9.	99
9a.	99
9b.	99
9c.	99
9d.	99
9e. Exercises	99

Chapter 10.	101
10a.	101
10b.	101
10c.	101
10d.	101
10e. Exercises	101
Chapter 11.	103
11a.	103
11b.	103
11c.	103
11d.	103
11e. Exercises	103
Chapter 12.	105
12a.	105
12b.	105
12c.	105
12d.	105
12e. Exercises	105
Part IV. Free electrodynamics	107
Chapter 13.	109
13a.	109
13b.	109
13c.	109
13d.	109
13e. Exercises	109
Chapter 14.	111
14a.	111
14b.	111
14c.	111
14d.	111
14e. Exercises	111
Chapter 15.	113

15a.	113
15b.	113
15c.	113
15d.	113
15e. Exercises	113
Chapter 16.	115
16a.	115
16b.	115
16c.	115
16d.	115
16e. Exercises	115
Bibliography	117

Part I

Classical electrodynamics

*Plain talking
Take us so far
Broken down cars
Like strung out old stars*

CHAPTER 1

Coulomb, Gauss

1a. Point charges

What is a charge? Not an easy question. The first thought goes to a magnet, or perhaps battery, but these have + and – ends, and so are something more complicated. The second thought goes to something like electricity, but that’s rather moving charges, electrons – travelling, and as explained above, too complicated, for later. As a third thought now, why not an electron – itself? But if we agree on this, we need a positive buddy for our electron, in our theory, and that can only be the proton +, and the thing now is that this couple electron/proton is exactly the hydrogen atom, rather belonging to quantum mechanics, and too complicated, again for later.

So, in the lack of anything simple, we have to start at a somewhat advanced level, physically speaking, but also very down-to-earth, just speaking like this, as follows:

FACT 1.1. Ordinary matter is made of electrons –, protons + and neutrons 0, with the number of + and – being rigorously equal, up to tiny tolerances. When the number of – is greater than the number of +, or vice versa, we say that we have a charge.

This is something quite interesting already, with the “tiny tolerances” mentioned above being, perhaps quite suprisingly, of the order of less than 10^{-10} . So when you touch a Van de Graaff generator, even after cranking well, please be sure that you won’t be a Terminator afterwards, but still well within that modest 10^{-10} tolerance. At $> 10^{-10}$ things violently explode, as bit as masses can explode too, due to $E = mc^2$.

In order to axiomatize our theory, we will proceed a bit like for gravity. We will assume that charges $q = \#p - \#e$ as in Fact 1.1 are no longer quantized, $q \in \mathbb{R}$, that they are points, and that they live in the void. Thus, we are led to:

DEFINITION 1.2. An electrostatic charge is a point $x \in \mathbb{R}^3$ having associated to it a certain number $q \in \mathbb{R}$, called charge of that point, and living in the void.

Here the last part, referring to the void, is something quite subtle, corresponding to a phenomenon not appearing in gravitation. In gravitation we know well about friction and drag, the two bad guys, but these affect the object itself, or rather its movement, and not the gravitation force which produces this movement. Things are not like this in electrodynamics, where matter in between objects affects the magnitude of the attraction

or repulsion force, even before it comes to movement, and with the explanation of this coming somehow from the picture of matter from Fact 1.1. Thus, we need void.

We have now all the needed ingredients for getting started, with:

FACT 1.3 (Coulomb law). *Any pair of charges $q_1, q_2 \in \mathbb{R}$ is subject to a force as follows, which is attractive if $q_1 q_2 < 0$ and repulsive if $q_1 q_2 > 0$,*

$$\|F\| = K \cdot \frac{|q_1 q_2|}{d^2}$$

where $d > 0$ is the distance between the charges, and $K > 0$ is a certain constant.

Observe the amazing similarity with the Newton law for gravity. However, as explained in the above, passed a few simple facts, things will be more complicated here.

As in the gravity case, the force F appearing above is understood to be parallel to the vector $x_2 - x_1 \in \mathbb{R}^3$ joining as $x_1 \rightarrow x_2$ the locations $x_1, x_2 \in \mathbb{R}^3$ of our charges, and by taking into account the attraction/repulsion rules above, we have:

PROPOSITION 1.4. *The Coulomb force of q_1 at x_1 acting on q_2 at x_2 is*

$$F = K \cdot \frac{q_1 q_2 (x_2 - x_1)}{\|x_2 - x_1\|^3}$$

with $K > 0$ being the Coulomb constant, as above.

PROOF. We have indeed the following computation:

$$\begin{aligned} F &= \operatorname{sgn}(q_1 q_2) \cdot \|F\| \cdot \frac{x_2 - x_1}{\|x_2 - x_1\|} \\ &= \operatorname{sgn}(q_1 q_2) \cdot K \cdot \frac{|q_1 q_2|}{\|x_2 - x_1\|^2} \cdot \frac{x_2 - x_1}{\|x_2 - x_1\|} \\ &= K \cdot \frac{q_1 q_2 (x_2 - x_1)}{\|x_2 - x_1\|^3} \end{aligned}$$

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

In relation now with the value of the constant K appearing in the above, called Coulomb constant, things here are a bit tricky, as follows:

FACT 1.5. *The Coulomb constant K is given by the formula*

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

in standard units, with the charges being measured in coulombs C , given by

$$1C \simeq 6.241\,509 \times 10^{18} e$$

where e is the elementary charge, namely minus that of an electron.

There are in fact several interesting things going on here. First, at the end you would say why not simply saying that e is the charge of the proton, but the thing is that the proton and the electron do not have in fact the same exact charge, with sign switched, and the electron was preferred, as always, over the proton for formulating things.

Which takes us into the question of why the charge of the electron is $-$, instead of $+$. And there is a long story here, involving debates among the 18th century greats, and with a little bit of confusion being involved too, because the electrons $-$ are attracted by positive charges $q > 0$, and so observed around these positive charges $q > 0$, which might lead to the idea that they might have themselves a positive charge $+$, contributing to $q > 0$. Benjamin Franklin is generally credited for the $-$ convention.

Things were later restored in the early 20th century, with the atomic theory of Bohr and others, where electrons $-$ spin around a proton and neutron core $q > 0$, and with this picture, including the signs, looking like something very reasonable.

Passed all this, another peculiarity of Fact 1.5 comes in relation with the definition of the coulomb, which is in fact given by definition by an exact formula, namely:

$$1C = \frac{5 \times 10^{18}}{0.801\ 088\ 317} e$$

This in practice gives the following more precise formula for the coulomb, which shows that a charge of $1C$ is something fractionary, that cannot be realized in real life:

$$1C = 6241\ 509\ 074\ 460\ 762\ 607.776 e$$

The problem comes from the following alternative definition of the coulomb, in terms of the ampere, which is something more complicated, that we will talk about later:

$$1C = 1A \cdot 1s$$

Hang on, we are not done yet. Adding to the confusion, the Coulomb constant is usually denoted K , but also k , or most often k_e , but in fact the most often is written in the following form, with ε_0 being the so-called permittivity of free space:

$$K = \frac{1}{4\pi\varepsilon_0}$$

And the story is not over here, because ε_0 itself is given by the following formula, with μ_0 being the magnetic permeability of free space, and c being the speed of light:

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

And we are surely still not done, because all the above discussion assumes that the other units that are used are standard, namely meter and second, and this is not always standard, due to the about 50 orders of magnitude physics has to deal with.

In any case, let us end this interesting discussion about units with something concrete, useful, and very illustrating, in relation with gravity, as follows:

THEOREM 1.6. *The electrical repulsion between two electrons is about*

$$R = 10^{42}$$

times bigger than their gravitational attraction.

PROOF. Consider indeed two electrons, having masses m, m and charges $-e, -e$. The magnitudes of the electric repulsion F_e and gravity attraction F_g are given by:

$$\|F_e\| = \frac{Ke^2}{d^2} \quad , \quad \|F_g\| = \frac{Gm^2}{d^2}$$

Thus the ratio of forces R that we want to measure is given by:

$$R = \frac{\|F_e\|}{\|F_g\|} = \frac{Ke^2}{Gm^2}$$

Regarding now the data, this is as follows, with m at rest, and in standard units, namely meters and seconds, also kilograms, and including now coulombs too:

$$K = 8.897 \times 10^9$$

$$G = 6.674 \times 10^{-11}$$

$$e = 1.602 \times 10^{-19}$$

$$m = 9.109 \times 10^{-31}$$

We obtain the following approximation for the ratio R considered above:

$$\begin{aligned} R &= \frac{8.897 \times 1.602^2}{6.674 \times 9.109^2} \times \frac{10^9 \times 10^{-38}}{10^{-11} \times 10^{-62}} \\ &= (4.123 \times 10^{-2}) \times 10^{44} \\ &\simeq 10^{42} \end{aligned}$$

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

For adding to the picture, and in order to fully understand what that $R = 10^{42}$ number that we found truly means, let us complement the above result with:

PROPOSITION 1.7. *The universe, or at least the known universe, is about*

$$r = 10^{37}$$

bigger than a hydrogen atom, with this ratio being 10,000 smaller than R .

PROOF. The radius of the hydrogen atom can be anywhere between 25–120 pm, with $1 \text{ pm} = 10^{-12} \text{ m}$, depending on the convention used, with a commonly accepted figure being 53 pm, representing the mean distance between the proton and the electron. As for the radius of the known universe, again there is a story here, with a commonly accepted figure being $4.4 \times 10^{26} \text{ m}$. Thus the ratio that we are interested in is:

$$r = \frac{4.4 \times 10^{26}}{53 \times 10^{-12}} \simeq 10^{37}$$

And this is 10,000 smaller than 10^{42} , as claimed. \square

As a side comment, however, when speaking masses instead of sizes, the number $R = 10^{42}$ pales when compared to the mass of the known universe, counting ordinary mass only, accounting for 4.9%, divided by the mass of a hydrogen atom, which is:

$$\mathfrak{R} = \frac{1.5 \times 10^{53}}{1.8 \times 10^{-30}} \simeq 10^{83}$$

Getting back now to Theorem 1.6 as it is, let us point out that this is something not at all anecdotal, even in the context of the most abstract theoretical physics that you can ever imagine, not to say pure mathematics, because of the following rule of thumb, which is something widely agreed upon, by most of the scientists:

RULE 1.8. *Don't ever expect the mathematics and physics to be the same, over 10 orders of magnitude or so.*

In other words, with this in hand, Theorem 1.6 tells us a very interesting thing, namely that the mathematics and physics of the Coulomb force $F_e \sim 1/d^2$ will be in fact very different from the mathematics and physics of the Newton force $F_g \sim 1/d^2$. We will see in what follows that indeed it is so, but it is of course far better to be warned in advance of the potential difficulties on the way. So, Theorem 1.6 is something very smart.

As a final comment, as already mentioned, Rule 1.8 is something widely agreed upon, by applied mathematicians and physicists, chemists and engineers, and with that 10 orders of magnitude being usually replaced by something far sharper, of type 2-3.

1b. Field lines

We develop now the basic math needed for electrostatics. Let us start with:

DEFINITION 1.9. *Given charges $q_1, \dots, q_k \in \mathbb{R}$ located at positions $x_1, \dots, x_k \in \mathbb{R}^3$, we define their electric field to be the vector function*

$$E(x) = K \sum_i \frac{q_i(x - x_i)}{\|x - x_i\|^3}$$

so that their force applied to a charge $Q \in \mathbb{R}$ positioned at $x \in \mathbb{R}^3$ is given by $F = QE$.

Observe the analogy with gravity, save for the fact that instead of masses $m > 0$ we have now charges $q \in \mathbb{R}$, and that at the level of constants, G gets replaced by K .

More generally, we will be interested in electric fields of various non-discrete configurations of charges, such as charged curves, surfaces and solid bodies. You already know about such things from classical mechanics, in the gravitational context, but the discussion there, involving the gravitational force of a solid body having non-trivial shape or density, is something rather specialized. In the electricity context, however, things like wires or metal sheets or solid bodies coming in all sorts of shapes, tailored for their purpose, play a key role, so this extension is essential. So, let us go ahead with:

DEFINITION 1.10. *The electric field of a charge configuration $L \subset \mathbb{R}^3$, with charge density function $\rho : L \rightarrow \mathbb{R}$, is the vector function*

$$E(x) = K \int_L \frac{\rho(z)(x - z)}{\|x - z\|^3} dz$$

so that the force of L applied to a charge Q positioned at x is given by $F = QE$.

With the above definitions in hand, it is most convenient now to forget about the charges, and focus on the study of the corresponding electric fields E .

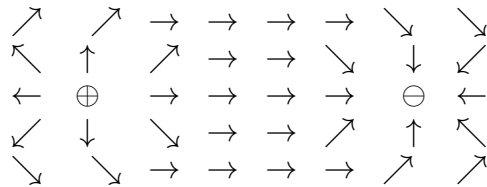
These fields are by definition vector functions $E : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, with the convention that they take $\pm\infty$ values at the places where the charges are located, and intuitively, are best represented by their field lines, which are constructed as follows:

DEFINITION 1.11. *The field lines of an electric field $E : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ are the oriented curves $\gamma \subset \mathbb{R}^3$ pointing at every point $x \in \mathbb{R}^3$ at the direction of the field, $E(x) \in \mathbb{R}^3$.*

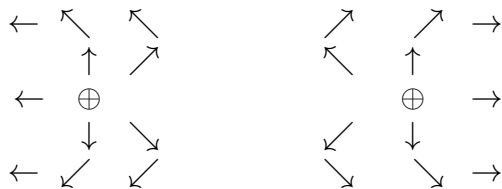
As a basic example here, for one charge the field lines are the half-lines emanating from its position, oriented according to the sign of the charge:



For two charges now, if these are of opposite signs, $+$ and $-$, you get a picture that you are very familiar with, namely that of the field lines of a bar magnet:



If the charges are $+, +$ or $-, -$, you get something of similar type, but repulsive this time, with the field lines emanating from the charges being no longer shared:



These pictures, and notably the last one, with $+, +$ charges, are quite interesting, because the repulsion situation does not appear in the context of gravity. Thus, we can only expect our geometry here to be far more complicated than that of gravity.

In general now, the first thing that can be said about the field lines is that, by definition, they do not cross. Thus, what we have here is some sort of oriented 1D foliation of \mathbb{R}^3 , in the sense that \mathbb{R}^3 is smoothly decomposed into oriented curves $\gamma \subset \mathbb{R}^3$.

The field lines, as constructed in Definition 1.11, obviously do not encapsulate the whole information about the field, with the direction of each vector $E(x) \in \mathbb{R}^3$ being there, but with the magnitude $\|E(x)\| \geq 0$ of this vector missing. However, say when drawing, when picking up uniformly radially spaced field lines around each charge, and with the number of these lines proportional to the magnitude of the charge, and then completing the picture, the density of the field lines around each point $x \in \mathbb{R}^3$ will give you then the magnitude $\|E(x)\| \geq 0$ of the field there, up to a scalar.

Let us summarize these observations as follows:

PROPOSITION 1.12. *Given an electric field $E : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, the knowledge of its field lines is the same as the knowledge of the composition*

$$nE : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \rightarrow S$$

where $S \subset \mathbb{R}^3$ is the unit sphere, and $n : \mathbb{R}^3 \rightarrow S$ is the rescaling map, namely:

$$n(x) = \frac{x}{\|x\|}$$

However, in practice, when the field lines are accurately drawn, the density of the field lines gives you the magnitude of the field, up to a scalar.

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

(1) The first assertion is clear from definitions, with of course our usual convention that the electric field and its problematics take place outside the locations of the charges, which makes everything in the statement to be indeed well-defined.

(2) Regarding now the last assertion, which is of course a bit informal, this follows from the above discussion. It is possible to be a bit more mathematical here, with a definition, formula and everything, but we will not need this, in what follows. \square

1c. The flux

Let us introduce now a key definition, as follows:

DEFINITION 1.13. *The flux of an electric field $E : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ through a surface $S \subset \mathbb{R}^3$, assumed to be oriented, is the quantity*

$$\Phi_E(S) = \int_S \langle E(x), n(x) \rangle dx$$

with $n(x)$ being unit vectors orthogonal to S , following the orientation of S . Intuitively, the flux measures the signed number of field lines crossing S .

Here by orientation of S we mean precisely the choice of unit vectors $n(x)$ as above, orthogonal to S , which must vary continuously with x . For instance a sphere has two possible orientations, one with all these vectors $n(x)$ pointing inside, and one with all these vectors $n(x)$ pointing outside. More generally, any surface has locally two possible orientations, so if it is connected, it has two possible orientations. In what follows the convention is that the closed surfaces are oriented with each $n(x)$ pointing outside.

Regarding the last sentence of Definition 1.13, this is of course something informal, meant to help, coming from the interpretation of the field lines from Proposition 1.12. However, we will see later that this simple interpretation can be of great use.

As a first illustration, let us do a basic computation, as follows:

PROPOSITION 1.14. *For a point charge $q \in \mathbb{R}$ at the center of a sphere S ,*

$$\Phi_E(S) = \frac{q}{\varepsilon_0}$$

where the constant is $\varepsilon_0 = 1/(4\pi K)$, independently of the radius of S .

PROOF. Assuming that S has radius r , we have the following computation:

$$\begin{aligned} \Phi_E(S) &= \int_S \langle E(x), n(x) \rangle dx \\ &= \int_S \left\langle \frac{Kqx}{r^3}, \frac{x}{r} \right\rangle dx \\ &= \int_S \frac{Kq}{r^2} dx \\ &= \frac{Kq}{r^2} \times 4\pi r^2 \\ &= 4\pi Kq \end{aligned}$$

Thus with $\varepsilon_0 = 1/(4\pi K)$ as above, we obtain the result. \square

More generally now, we have the following result:

THEOREM 1.15. *The flux of a field E through a sphere S is given by*

$$\Phi_E(S) = \frac{Q_{enc}}{\varepsilon_0}$$

where Q_{enc} is the total charge enclosed by S , and $\varepsilon_0 = 1/(4\pi K)$.

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

(1) Before jumping into computations, let us do some manipulations. First, by discretizing the problem, we can assume that we are dealing with a system of point charges. Moreover, by additivity, we can assume that we are dealing with a single charge. And if we denote by $q \in \mathbb{R}$ this charge, located at $v \in \mathbb{R}^3$, we want to prove that we have the following formula, where $B \subset \mathbb{R}^3$ denotes the ball enclosed by S :

$$\Phi_E(S) = \frac{q}{\varepsilon_0} \delta_{v \in B}$$

(2) By linearity we can assume that we are dealing with the unit sphere S . Moreover, by rotating we can assume that our charge q lies on the Ox axis, that is, that we have $v = (r, 0, 0)$ with $r \geq 0$, $r \neq 1$. The formula that we want to prove becomes:

$$\Phi_E(S) = \frac{q}{\varepsilon_0} \delta_{r < 1}$$

(3) Let us start now the computation. With $u = (x, y, z)$, we have:

$$\begin{aligned} \Phi_E(S) &= \int_S \langle E(u), u \rangle du \\ &= \int_S \left\langle \frac{Kq(u-v)}{\|u-v\|^3}, u \right\rangle du \\ &= Kq \int_S \frac{\langle u-v, u \rangle}{\|u-v\|^3} du \\ &= Kq \int_S \frac{1 - \langle v, u \rangle}{\|u-v\|^3} du \\ &= Kq \int_S \frac{1 - rx}{(1 - 2xr + r^2)^{3/2}} du \end{aligned}$$

(4) In order to compute the above integral, we will use spherical coordinates for the unit sphere S , which are as follows, with $s \in [0, \pi]$ and $t \in [0, 2\pi]$:

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

The corresponding Jacobian is readily computed, as follows:

$$\begin{aligned}
J &= \begin{vmatrix} \cos s & -\sin s & 0 \\ \sin s \cos t & \cos s \cos t & -\sin s \sin t \\ \sin s \sin t & \cos s \sin t & \sin s \cos t \end{vmatrix} \\
&= \sin s \sin t \begin{vmatrix} \cos s & -\sin s \\ \sin s \sin t & \cos s \sin t \end{vmatrix} + \sin s \cos t \begin{vmatrix} \cos s & -\sin s \\ \sin s \cos t & \cos s \cos t \end{vmatrix} \\
&= \sin s (\sin^2 t + \cos^2 t) \begin{vmatrix} \cos s & -\sin s \\ \sin s & \cos s \end{vmatrix} \\
&= \sin s
\end{aligned}$$

(5) With the above change of coordinates, our integral from (3) becomes:

$$\begin{aligned}
\Phi_E(S) &= Kq \int_S \frac{1 - rx}{(1 - 2rx + r^2)^{3/2}} du \\
&= Kq \int_0^{2\pi} \int_0^\pi \frac{1 - r \cos s}{(1 - 2r \cos s + r^2)^{3/2}} \cdot \sin s \, ds \, dt \\
&= 2\pi Kq \int_0^\pi \frac{(1 - r \cos s) \sin s}{(1 - 2r \cos s + r^2)^{3/2}} ds \\
&= \frac{q}{2\epsilon_0} \int_0^\pi \frac{(1 - r \cos s) \sin s}{(1 - 2r \cos s + r^2)^{3/2}} ds
\end{aligned}$$

(6) The point now is that the integral on the right can be computed with the change of variables $x = \cos s$. Indeed, we have $dx = -\sin s \, ds$, and we obtain:

$$\begin{aligned}
\int_0^\pi \frac{(1 - r \cos s) \sin s}{(1 - 2r \cos s + r^2)^{3/2}} ds &= \int_{-1}^1 \frac{1 - rx}{(1 - 2rx + r^2)^{3/2}} dx \\
&= \left[\frac{x - r}{\sqrt{1 - 2rx + r^2}} \right]_{-1}^1 \\
&= \frac{1 - r}{\sqrt{1 - 2r + r^2}} - \frac{-1 - r}{\sqrt{1 + 2r + r^2}} \\
&= \frac{1 - r}{|1 - r|} + 1 \\
&= 2\delta_{r < 1}
\end{aligned}$$

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

As a technical comment here, at $r = 1$, which is normally avoided by our problematics, the integral I_r computed in (6) above converges too, and can be evaluated as follows:

$$I_1 = \left[\frac{x-1}{\sqrt{2-2x}} \right]_{-1}^1 = \left[-\sqrt{\frac{1-x}{2}} \right]_{-1}^1 = 1$$

Thus, we have the correct middle step between the 0, 2 values of the integral I_r , and getting back now to the flux, at $r = 1$ we formally have $\Phi_E(S) = q/(2\varepsilon_0)$, which again is the correct middle step between the 0, q/ε_0 values of the flux.

1d. Gauss law

Even more generally now, we have the following result, due to Gauss, which is the foundation of advanced electrostatics, and of everything following from it, namely electrodynamics, and then quantum mechanics, and particle physics:

THEOREM 1.16 (Gauss law). *The flux of a field E through a surface S is given by*

$$\Phi_E(S) = \frac{Q_{enc}}{\varepsilon_0}$$

where Q_{enc} is the total charge enclosed by S , and $\varepsilon_0 = 1/(4\pi K)$.

PROOF. This basically follows from Theorem 1.15, or even from Proposition 1.14, by adding to the results there a number of new ingredients, as follows:

(1) Our first claim is that given a closed surface S , with no charges inside, the flux through it of any choice of external charges vanishes:

$$\Phi_E(S) = 0$$

This claim is indeed supported by the intuitive interpretation of the flux, as corresponding to the signed number of field lines crossing S . Indeed, any field line entering as $+$ must exit somewhere as $-$, and vice versa, so when summing we get 0.

(2) In practice now, in order to prove this rigorously, there are several ways. A first argument, which is quite elementary, is the one used by Feynman in [38], based on the fact that, due to $F \sim 1/d^2$, local deformations of S will leave invariant the flux, and so in the end we are left with a rotationally invariant surface, where the result is clear.

(3) A second argument, which basically uses the same idea, but is perhaps a bit more robust, is by redoing the computations in the proof of Theorem 1.15, by assuming this time that the integration takes place on an arbitrary surface as follows:

$$S_\lambda = \left\{ \lambda(u)u \mid u \in S \right\}$$

To be more precise, here $\lambda : S \rightarrow (0, \infty)$ is a certain function, defining the surface, whose derivatives will appear both in the construction of the normal vectors $n(x)$ with

$x = \lambda(u)u$, and in the Jacobian of the change of variables $x \rightarrow u$, and in the end, when integrating over S as in the proof of Theorem 1.15, this function λ disappears.

(4) A third argument, used by basically all electrodynamics books at the graduate level, and by some undergraduate books too, is by using heavy calculus, namely partial integration in 3D, and we will discuss this later, more in detail, in the next chapter.

(5) A fourth argument is by following the nice idea in (1), namely carefully axiomatizing the field lines, and their relation with the field, and then obtaining $\Phi_E(S) = 0$ by using the in-and-out trick in (1), as explained for instance by Griffiths in [44]. However, when looking for full rigor here, in practice this is something quite complicated, amounting more or less in proving the heavy 3D calculus results mentioned in (4) above via foliation methods, and we will not get here into this.

(6) To summarize, we are led to the conclusion that given a closed surface S , with no charges inside, the flux through it of any choice of external charges vanishes:

$$\Phi_E(S) = 0$$

(7) The point now is that, with this and Proposition 1.14 in hand, we can finish by using a standard math trick. Let us assume indeed, by discretizing, that our system of charges is discrete, consisting of enclosed charges $q_1, \dots, q_k \in \mathbb{R}$, and an exterior total charge Q_{ext} . We can surround each of q_1, \dots, q_k by small disjoint spheres U_1, \dots, U_k , chosen such that their interiors do not touch S , and we have:

$$\begin{aligned} \Phi_E(S) &= \Phi_E(S - \cup U_i) + \Phi_E(\cup U_i) \\ &= 0 + \Phi_E(\cup U_i) \\ &= \sum_i \Phi_E(U_i) \\ &= \sum_i \frac{q_i}{\varepsilon_0} \\ &= \frac{Q_{enc}}{\varepsilon_0} \end{aligned}$$

(8) To be more precise, in the above the union $\cup U_i$ is a usual disjoint union, and the flux is of course additive over components. As for the difference $S - \cup U_i$, this is by definition the disjoint union of S with the disjoint union $\cup(-U_i)$, with each $-U_i$ standing for U_i with orientation reversed, and since this difference has no enclosed charges, the flux through it vanishes by (6). Finally, the end makes use of Proposition 1.14. \square

1e. Exercises

Exercises.

CHAPTER 2

Poisson and Laplace

2a. Some calculus

We have seen that the study of electrostatics leads us into some serious 3D geometry. So, time to review and improve our math tools, for dealing with all this. To be more precise, we would like to extend our basic 1D calculus tools to 3D.

It is convenient to keep the discussion, when possible, in N dimensions, among others for covering at the same time 1D, 2D, 3D, and also 4D needed for relativity. Regarding the derivative, this is something that you know well, as follows:

PROPOSITION 2.1. *The derivative of a function $f : \mathbb{R}^N \rightarrow \mathbb{R}^M$ at a point $c \in \mathbb{R}^N$ is the linear map $f'(c) : \mathbb{R}^N \rightarrow \mathbb{R}^M$ which best approximates f around c ,*

$$f(c + z) \simeq f(c) + f'(c)z$$

regarded as a rectangular matrix, $f'(c) \in M_{M \times N}(\mathbb{R})$, acting on the vectors $z \in \mathbb{R}^N$, written vertically. With these conventions, the chain rule for derivatives holds,

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c)$$

in the sense that the matrix on the left is the product of the matrices on the right.

PROOF. This is something that we are very familiar with, and the proof, which is elementary, can be found in any multivariable calculus book. \square

Next in line, we have the formula of change of variables. Here the result, that we already used in the proof of the Gauss law for the sphere, is as follows:

THEOREM 2.2. *Given a function $f : E \rightarrow \mathbb{R}$, with $E \subset \mathbb{R}^N$, and a transformation in N variables, $\varphi = (\varphi_1, \dots, \varphi_N)$, we have the change of variable formula,*

$$\int_E f(x)dx = \int_{\varphi^{-1}(E)} f(\varphi(y))|J_\varphi(y)|dy$$

with the J_φ quantity on the right, called Jacobian of φ , being given by:

$$J_\varphi(y) = \det \left[\left(\frac{d\varphi_i}{dx_j}(y) \right)_{ij} \right]$$

For spherical coordinates, the Jacobian is $J_\varphi = r^{N-1}$ times a product of sines.

PROOF. Again, this is something that you are familiar with, as follows:

(1) When performing a change of variables $\varphi : \mathbb{R}^N \rightarrow \mathbb{R}^N$, the coefficient that appears when integrating is the “inflation coefficient” of φ , around the given point $y \in \mathbb{R}^N$. In 1D this is just the number $\varphi'(y)$, or rather its absolute value $|\varphi'(y)|$, when putting the bounds as above. In N dimensions, however, $\varphi'(y)$ is no longer a number, so what to do.

(2) In order to solve the problem, the first remark is that, according to Proposition 2.1, the function φ is best approximated around y by its derivative, as follows:

$$\varphi(y + u) = \varphi(y) + \varphi'(y)u$$

Moreover, by linear algebra, this derivative is a square matrix, given by:

$$\varphi'(y) = \left(\frac{d\varphi_i}{dx_j}(y) \right)_{ij}$$

(3) Now recall that the determinant of a square matrix A having column vectors $v_1, \dots, v_N \in \mathbb{R}^N$ is the signed volume of the parallelepiped formed by these vectors:

$$\det(v_1 \dots v_N) = \pm \text{vol} \langle v_1, \dots, v_N \rangle$$

Thus, the inflation coefficient of A is the number $|\det A| \geq 0$, and getting back now to our problem, $|J_\varphi(y)| = |\det \varphi'(y)|$ is the number that must be added, as claimed.

(4) Finally, in what regards spherical coordinates, these are as follows:

$$\begin{cases} x_1 &= r \cos t_1 \\ x_2 &= r \sin t_1 \cos t_2 \\ \vdots & \\ x_{N-1} &= r \sin t_1 \sin t_2 \dots \sin t_{N-2} \cos t_{N-1} \\ x_N &= r \sin t_1 \sin t_2 \dots \sin t_{N-2} \sin t_{N-1} \end{cases}$$

We are already familiar with this in 2D, 3D. In N dimensions now, the Jacobian can be computed exactly as in 2D, 3D, and is given by the following formula:

$$J = r^{N-1} \sin^{N-2} t_1 \sin^{N-3} t_2 \dots \sin^2 t_{N-3} \sin t_{N-2}$$

(5) So, this was for the idea, and the detailed proof of all the above, except for (4) which is easy, to be best done by yourself, can be found in any calculus book. \square

Moving ahead now, if there is one thing missing from our picture, this is the scary one, namely partial integration. Things here are quite tricky in N dimensions, but in the cases where we are mainly interested in, $N = 2, 3$, we have several useful results. Let us start with a standard definition, immersing us into 3D problematics, as follows:

DEFINITION 2.3. Given a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, its usual derivative $f'(u) \in \mathbb{R}^3$ can be written as $f'(u) = \nabla f(u)$, where the gradient operator ∇ is given by:

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

By using ∇ , we can talk about the divergence of a function $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, as being

$$\langle \nabla, \varphi \rangle = \left\langle \begin{pmatrix} \frac{d}{dx} \\ \frac{d}{dy} \\ \frac{d}{dz} \end{pmatrix}, \begin{pmatrix} \varphi_x \\ \varphi_y \\ \varphi_z \end{pmatrix} \right\rangle = \frac{d\varphi_x}{dx} + \frac{d\varphi_y}{dy} + \frac{d\varphi_z}{dz}$$

as well as about the curl of the same function $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, as being

$$\nabla \times \varphi = \begin{vmatrix} u_x & \frac{d}{dx} & \varphi_x \\ u_y & \frac{d}{dy} & \varphi_y \\ u_z & \frac{d}{dz} & \varphi_z \end{vmatrix} = \begin{pmatrix} \frac{d\varphi_z}{dy} - \frac{d\varphi_y}{dz} \\ \frac{d\varphi_x}{dz} - \frac{d\varphi_z}{dx} \\ \frac{d\varphi_y}{dx} - \frac{d\varphi_x}{dy} \end{pmatrix}$$

where u_x, u_y, u_z are the unit vectors along the coordinate directions x, y, z .

Getting back now to calculus tools, what was missing from our picture was the higher dimensional analogue of the fundamental theorem of calculus, and more generally of the partial integration formula. In 3 dimensions, we have the following result:

THEOREM 2.4. The following results hold, in 3 dimensions:

- (1) Fundamental theorem for gradients, namely

$$\int_a^b \langle \nabla f, dx \rangle = f(b) - f(a)$$

- (2) Fundamental theorem for divergences, or Gauss or Green formula,

$$\int_B \langle \nabla, \varphi \rangle = \int_S \langle \varphi(x), n(x) \rangle dx$$

- (3) Fundamental theorem for curls, or Stokes formula,

$$\int_A \langle (\nabla \times \varphi)(x), n(x) \rangle dx = \int_P \langle \varphi(x), dx \rangle$$

where S is the boundary of the body B , and P is the boundary of the area A .

PROOF. This is a mixture of trivial and non-trivial results, as follows:

- (1) This is something that we know well in 1D, namely the fundamental theorem of calculus, and the general, N -dimensional formula follows from that.

(2) This is something more subtle, and we had a taste of it when dealing with the Gauss law, and its various proofs. In general, the proof is similar, by using the various ideas from the proof of the Gauss law, and this can be found in any calculus book.

(3) This is again something subtle, and again with a flavor of things that we know, from the proof of the Gauss law, and which can be found in any calculus book. \square

2b. Gauss, again

Getting back now to electrostatics, as a first application of the above, we have the following new point of view on the Gauss formula, which is more conceptual:

THEOREM 2.5 (Gauss). *Given an electric potential E , its divergence is given by*

$$\langle \nabla, E \rangle = \frac{\rho}{\varepsilon_0}$$

where ρ denotes as usual the charge distribution. Also, we have

$$\nabla \times E = 0$$

meaning that the curl of E vanishes.

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

(1) The first formula, called Gauss law in differential form, follows from:

$$\begin{aligned} \int_B \langle \nabla, E \rangle &= \int_S \langle E(x), n(x) \rangle dx \\ &= \Phi_E(S) \\ &= \frac{Q_{enc}}{\varepsilon_0} \\ &= \int_B \frac{\rho}{\varepsilon_0} \end{aligned}$$

Now since this must hold for any B , this gives the formula in the statement.

(2) As a side remark, the Gauss law in differential form can be established as well directly, with the computation, involving a Dirac mass, being as follows:

$$\begin{aligned}
 \langle \nabla, E \rangle (x) &= \left\langle \nabla, K \int_{\mathbb{R}^3} \frac{\rho(z)(x-z)}{\|x-z\|^3} dz \right\rangle \\
 &= K \int_{\mathbb{R}^3} \left\langle \nabla, \frac{x-z}{\|x-z\|^3} \right\rangle \rho(z) dz \\
 &= K \int_{\mathbb{R}^3} 4\pi \delta_x \cdot \rho(z) dz \\
 &= 4\pi K \int_{\mathbb{R}^3} \delta_x \rho(z) dz \\
 &= \frac{\rho(x)}{\varepsilon_0}
 \end{aligned}$$

And with this in hand, we have via (1) a new proof of the usual Gauss law.

(3) Regarding the curl, by discretizing and linearity we can assume that we are dealing with a single charge q , positioned at 0. We have, by using spherical coordinates r, s, t :

$$\begin{aligned}
 \int_a^b \langle E(x), dx \rangle &= \int_a^b \left\langle \frac{Kqx}{\|x\|^3}, dx \right\rangle \\
 &= \int_a^b \left\langle \frac{Kq}{r^2} \cdot \frac{x}{\|x\|}, dx \right\rangle \\
 &= \int_a^b \frac{Kq}{r^2} dr \\
 &= \left[-\frac{Kq}{r} \right]_a^b \\
 &= Kq \left(\frac{1}{r_a} - \frac{1}{r_b} \right)
 \end{aligned}$$

In particular the integral of E over any closed loop vanishes, and by using now Stokes' theorem, we conclude that the curl of E vanishes, as stated.

(4) Finally, as a side remark, both the formula of the divergence and the vanishing of the curl are somewhat clear by looking at the field lines of E . However, as all the above mathematics shows, there is certainly something to be understood, in all this. \square

2c. Poisson equation

With this done, let us discuss now energy and potentials. Recall from classical mechanics the formula $F = -\nabla V$? The same holds in the present setting. We first have:

THEOREM 2.6. *Consider an electric field, given as usual by:*

$$E(x) = K \int_L \frac{\rho(z)(x-z)}{\|x-z\|^3} dz$$

We have then $E = -\nabla V$, with the corresponding potential V being given by

$$V = K \int_L \frac{\rho(z)}{\|x-z\|} dz$$

and the usual work and energy considerations for conservative forces hold.

PROOF. Generally speaking, all this is something that you know well from classical mechanics. However, there are a few notable differences here, as follows:

(1) First of all, strange things happen when allowing charges to move, and we don't know yet about that, so such energy considerations remain something quite formal.

(2) In what regards the formula for V in the statement, this is the usual formula for gravity, with masses replaced by charges, and with two other changes, as follows:

(3) First, the previous $-$ sign from gravity has disappeared, because in the gravitational context $m_1 m_2 > 0$, always true, corresponds to an attractive force, while in our setting $q_1 q_2 > 0$ corresponds to a repulsive force. Thus, we must change the sign.

(4) And second, as already mentioned in (1), things here are a bit formal, so we have chosen to divide V , previously in gravity thought as being potential energy, by the receiving charge, as for this V to be a feature of the electric field E only. \square

Moving ahead now, the question appears, what happens to the Gauss equation for the electric field E , as formulated in chapter 1, when written in terms of the associated potential V . And the answer here, which is remarkable, is as follows:

THEOREM 2.7 (Poisson). *In terms of the electric potential V , the Gauss formula becomes the Poisson equation, namely*

$$\Delta V = -\frac{\rho}{\varepsilon_0}$$

with $\Delta = \langle \nabla, \nabla \rangle$ being the Laplace operator, given by the formula

$$\Delta f = \sum_i \frac{d^2 f}{dx_i^2}$$

and the curl equation disappears, being automatic for gradients.

PROOF. Here both the assertions are elementary, as follows:

(1) With $E = -\nabla V$ the Gauss equation $\langle \nabla, E \rangle = \rho/\varepsilon_0$ becomes:

$$\langle \nabla, \nabla V \rangle = -\frac{\rho}{\varepsilon_0}$$

Thus we must have $\Delta V = -\rho/\varepsilon_0$, with the operator Δ being given by:

$$\begin{aligned}\Delta f &= \langle \nabla, \nabla f \rangle \\ &= \left\langle \left(\frac{d}{dx} \right), \left(\frac{df}{dx} \right) \right\rangle \\ &= \frac{d^2 f}{dx^2} + \frac{d^2 f}{dy^2} + \frac{d^2 f}{dz^2}\end{aligned}$$

Thus, we are led to the Poisson equation in the statement.

(2) Regarding now the curl, our claim is that the equation $\nabla \times E = 0$ simply disappears, this type of vanishing being automatic for gradients. Indeed, for any f we have:

$$\begin{aligned}\nabla \times \nabla f &= \begin{vmatrix} u_x & \frac{d}{dx} & \frac{df}{dx} \\ u_y & \frac{d}{dy} & \frac{df}{dy} \\ u_z & \frac{d}{dz} & \frac{df}{dz} \end{vmatrix} \\ &= \begin{pmatrix} \frac{d^2 f}{dydz} - \frac{d^2 f}{dzdy} \\ \frac{d^2 f}{dzdx} - \frac{d^2 f}{dx dz} \\ \frac{d^2 f}{dx dy} - \frac{d^2 f}{dy dx} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\end{aligned}$$

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

As an interesting feature of the potential approach, the Poisson equation makes sense, and is in fact very interesting, even when no charge is present, and we have here:

THEOREM 2.8 (Laplace). *In the case where no charges are present, the Poisson equation, and so the Gauss and even Coulomb laws too, in a certain sense, become*

$$\Delta V = 0$$

called Laplace equation, whose solutions are called harmonic functions. These functions have an interesting mathematics, reminding that of the holomorphic functions in 2D.

PROOF. There are many things that can be said here, as follows:

(1) First of all, the Laplace equation and its physical meaning come from the Poisson equation, and from the various potential considerations in the above.

(2) Mathematically now, the idea is that various remarkable results about the holomorphic functions in 2D, such as the mean formula, extend to the harmonic functions. \square

2d. Faraday cage

As a main application of all this, let us discuss now the Faraday cage. You have probably heard of it, and maybe thought a bit at its validity too, when stuck during a huge rainstorm, in your metal car, in the hope that the Faraday cage principle will indeed work. We will see now, with proof, that the Faraday cage principle always works:

THEOREM 2.9 (Faraday cage). *In a cavity surrounded by a conductor, the electric field is 0. Moreover, the same principle holds when the cavity is not exactly perfect.*

PROOF. This is based on some more physics, the idea being as follows:

(1) There are all sorts of materials, but these basically fall into two classes. First we have the insulators, like plastic or glass, where the electrons are not free to move. And then we have the conductors, like metals, where there are electrons which can freely move. Conductors include as well other materials, at specific conditions and so on, but in what follows we will be interested in the “perfect conductors”, those having an unlimited supply of free electrons, and metals at normal conditions are close to this status.

(2) Let us first try to understand what happens when putting a charge $q > 0$ near a conductor C . Since electrons are free to move, they will tend to go to the side of C which is close to q , making that side a charge $-$. And on the opposite side, we will have a lack of electrons, accounting for a charge $+$, the picture being as follows:

$$\begin{array}{c} \ominus \oplus \\ \oplus_q \quad \ominus C \oplus \\ \ominus \oplus \end{array}$$

(3) The point now is that, for the above to happen, an induced charge on C , we don't really need a charge $q > 0$, but just an electric field E . To be more precise, when putting C in no matter which electric field E , what will happen is that the electrons inside C will start moving, as to eventually produce a charge at the surface of C , producing itself a field $-E$ inside the conductor, countering the external field E .

(4) In particular, when putting a conductor C in the path of an electric field E , no matter what surface phenomena will happen, as explained above, and which of course are in direct relation with the magnitude of E , the field inside the conductor will be always $E + (-E) = 0$. Thus, you are safe in a cavity inside C , as stated. \square

As a conclusion to the general theory developed in this chapter, you won't be electrocuted when inside a Faraday cage. Note however that you might get cooked.

2e. Exercises

Exercises.

CHAPTER 3

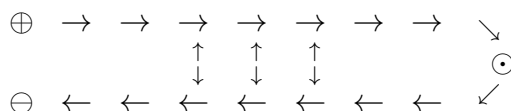
Maxwell equations

3a. Magnetism

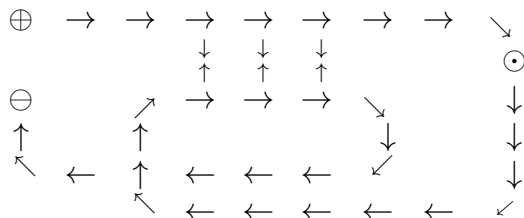
Magnetism, eventually. Who has not played with magnets as a kid, wishing to use them later at work, as an engineer, or at least understand them later, as a scientist. If you're reading now this book, magnets have probably something to do with it.

Unfortunately magnets are quite complicated things, and we will have to start our discussion with motors. And not even with cool gas engines, but with vulgar, electric motors. Yes, you heard me right, electric motors. So, let us do this, and we'll come back to magnets short after, that's promised. As for cool gas engines and Mad Max, we will come back to them too, in the next chapter, when discussing thermodynamics.

In order to understand the functioning of an electric motor, we don't need an actual motor, to start with, but just a battery feeding a light bulb. As a first observation, in a normal configuration of our device, the feeding cables will repel each other:



This is already quite surprising, and things are not over here. Indeed, when twisting a bit the cables, as to see what happens to parallel currents when moving in the same direction too, the conclusion is that in this case, we have attraction:

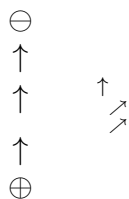


Summarizing, just by feeding a light bulb with a battery, and looking at the cables, and playing a bit with them, we are led to the following interesting conclusion:

FACT 3.1. *Parallel electric currents in opposite directions repel, and parallel electric currents in the same direction attract.*

We can in fact say even more, by further playing with the cables, armed this time with a compass. The conclusion is that each cable produces some kind of “magnetic field” around it, which interestingly, is not oriented in the direction of the current, but orthogonal to it, given by the right-hand rule, as follows:

FACT 3.2 (Right-hand rule). *An electric current produces a magnetic field B which is orthogonal to it, whose direction is given by the right-hand rule,*



namely wrap your right hand around the cable, with the thumb pointing towards the direction of the current, and the movement of your wrist will give you the direction of B .

This is something even more interesting than Fact 3.1. Indeed, not only moving charges produce something new, that we’ll have to investigate, but they know well about 3D, and more specifically about orientation there, left and right, even if living in 1D.

And isn’t this amazing. Let us summarize this discussion with:

FACT 3.3. *Charges are smart, they know about 3D, and about left and right.*

This invites to some philosophy, before moving ahead with some further physics, and then math. What is smartness? Not an easy question, but with our physics knowledge so far, we have at least two answers to it, as follows:

(1) Human smartness comes from chemical reactions in the brain, which reactions come from certain electrons taking certain decisions. But these electrons are exactly our charges, so human smartness ultimately comes from the charge smartness in Fact 3.3.

(2) Which leads us into the question whether masses are smart too. Not clear, but is there’s something to be said here, you have to agree that a tiny mass m exploding next to you, and releasing all its $E = mc^2$ energy, was probably smarter than you.

With this discussed, let us go ahead and investigate the charge smartness, and more specifically the magnetic fields discovered above.

In order to evaluate the properties of the magnetic fields B coming from electric currents, as in Fact 3.2, the simplest way is that of making them act on exterior charges Q . And we have here the following formula, to start with, due to Lorentz:

FACT 3.4 (Lorentz force law). *The magnetic force on a charge Q , moving with velocity v in a magnetic field B , is as follows, with \times being a vector product:*

$$F_m = (v \times B)Q$$

In the presence of both electric and magnetic fields, the total force on Q is

$$F = (E + v \times B)Q$$

where E is the electric field.

Here the occurrence of the vector product \times is not surprising, due to the fact that the right-hand rule appears both in Fact 3.2, and in the definition of \times . In fact, the Lorentz force law is just a fancy mathematical reformulation of Fact 3.2, telling us that, once the magnetic fields B duly axiomatized, and with this being a remaining big problem, their action on exterior charges Q will be proportional to the charge, $F_m \sim Q$, and with the orientation and magnitude coming from the 3D of the right-hand rule in Fact 3.2.

As an interesting application of the Lorentz force law, we have:

THEOREM 3.5. *Magnetic forces do not work.*

PROOF. This might seem quite surprising, but the math is there, as follows:

$$\begin{aligned} dW_m &= \langle F_m, dx \rangle \\ &= \langle (v \times B)Q, v dt \rangle \\ &= Q \langle v \times B, v \rangle dt \\ &= 0 \end{aligned}$$

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

3b. Electric currents

Moving ahead now, let us talk axiomatization of electric currents, including units. We have here the following definition, clarifying our previous discussion about coulombs:

DEFINITION 3.6. *The electric currents I are measured in amperes, given by:*

$$1A = 1C/s$$

As a consequence, the coulomb is given by $1C = 1A \times 1s$.

With this notion in hand, let us keep building the math and physics of magnetism. So, assume that we are dealing with an electric current I , producing a magnetic field B . In this context, the Lorentz force law from Fact 3.4 takes the following form:

$$F_m = \int (dx \times B)I$$

The current being typically constant along the wire, this reads:

$$F_m = I \int dx \times B$$

We can deduce from this the following result:

THEOREM 3.7. *The volume current density J satisfies*

$$\langle \nabla, J \rangle = -\dot{\rho}$$

called continuity equation.

PROOF. We have indeed the following computation, for any surface S enclosing a volume V , based on the Lorentz force law, and on the overall charge conservation:

$$\begin{aligned} \int_V \langle \nabla, J \rangle &= \int_S \langle J, n(x) \rangle dx \\ &= -\frac{d}{dt} \int_V \rho \\ &= -\int_V \dot{\rho} \end{aligned}$$

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

Moving ahead now, let us formulate the following definition:

DEFINITION 3.8. *The realm of magnetostatics is that of the steady currents,*

$$\dot{\rho} = 0 \quad , \quad \dot{J} = 0$$

in analogy with electrostatics, dealing with fixed charges.

As a first observation, for steady currents the continuity equation reads:

$$\langle \nabla, J \rangle = 0$$

We have here a bit of analogy between electrostatics and magnetostatics, and with this in mind, let us look for equations for the magnetic field B . We have:

FACT 3.9 (Biot-Savart law). *The magnetic field of a steady line current is given by*

$$B = \frac{\mu_0}{4\pi} \int \frac{I \times x}{||x||^3}$$

where μ_0 is a certain constant, called the magnetic permeability of free space.

This law not only gives us all we need, for studying steady currents, and we will talk about this in a moment, with math and everything, but also makes an amazing link with the Coulomb force law, due to the following fact, which is also part of it:

FACT 3.10 (Biot-Savart, continued). *The electric permittivity of free space ε_0 and the magnetic permeability of free space μ_0 are related by the formula*

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

where c is as usual the speed of light.

This is something truly remarkable, and very deep, that will have numerous consequences, in what follows, be that for investigating phenomena like radiation, or for making the link with Einstein's relativity theory, both crucially involving c .

But, first of all, this is certainly an invitation to rediscuss units and constants, as a continuation of our discussion from chapter 1. In what regards the units, we won't be impressed by the ampere, and keep using the coulomb, as a main unit:

CONVENTIONS 3.11. *We keep using standard units, namely meters, kilograms, seconds, along with the coulomb, defined by the following exact formula*

$$1C = \frac{5 \times 10^{18}}{0.801\,088\,317} e$$

with e being minus the charge of the electron, which in practice means:

$$1C \simeq 6.241 \times 10^{18} e$$

We will also use the ampere, defined as $1A = 1C/s$, for measuring currents.

In what regards constants, however, time to do some cleanup. We have been boycotting for some time already the Coulomb constant K , and using instead $\varepsilon_0 = 1/(4\pi K)$, due to the ubiquitous 4π factor, first appearing as the area of the unit sphere, $A = 4\pi$, in the computation for the Gauss law for the unit sphere. Together with Fact 3.10, this suggests using the numbers ε_0, μ_0 as our new constants, by always keeping in mind $\varepsilon_0\mu_0 = 1/c^2$, and by having of course c as constant too, and we are led in this way into:

CONVENTIONS 3.12. *We use from now on as constants the electric permittivity of free space ε_0 and the magnetic permeability of free space μ_0 , given by*

$$\varepsilon_0 = 8.854\,187\,8128(13) \times 10^{-12}$$

$$\mu_0 = 1.256\,637\,062\,12(19) \times 10^{-6}$$

as well as the speed of light, given by the following exact formula,

$$c = 299\,792\,458$$

which are related by $\varepsilon_0\mu_0 = 1/c^2$, and with the Coulomb constant being $K = 1/(4\pi\varepsilon_0)$.

Observe in passing that we are not messing up our figures, which can be quite often the case in this type of situation, because according to our data, and by truncating instead of rounding, as busy theoretical physicists usually do, we have:

$$\varepsilon_0 \mu_0 c^2 = 8.854 \times 1.256 \times 2.997^2 \times 10^{16-12-6} = 0.998$$

We will be back to numerics later on, when rediscussing as well the relation between electricity and gravitation, following a discussion from chapter 1, but this time with magnetism added. As an obvious comment, however, observe how the above figures show that magnetic forces are far weaker than electric forces. More on this later.

Getting back now to theory and math, the Biot-Savart law has as consequence:

THEOREM 3.13. *We have the following formula:*

$$\langle \nabla, B \rangle = 0$$

PROOF. We recall that the Biot-Savart law tells us that the magnetic field B of a steady line current I is given by the following formula:

$$B = \frac{\mu_0}{4\pi} \int \frac{I \times x}{\|x\|^3}$$

By applying the divergence operator to this formula, we obtain:

$$\begin{aligned} \langle \nabla, B \rangle &= \frac{\mu_0}{4\pi} \int \left\langle \nabla, \frac{I \times x}{\|x\|^3} \right\rangle \\ &= \frac{\mu_0}{4\pi} \int \left\langle \nabla \times J, \frac{x}{\|x\|^3} \right\rangle - \left\langle \nabla \times \frac{x}{\|x\|^3}, J \right\rangle \\ &= \frac{\mu_0}{4\pi} \int \left\langle 0, \frac{x}{\|x\|^3} \right\rangle - \langle 0, J \rangle \\ &= 0 \end{aligned}$$

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

Regarding now the curl, we have here a similar result, as follows:

THEOREM 3.14 (Ampère law). *We have the following formula:*

$$\nabla \times B = \mu_0 J$$

PROOF. Again, we use the Biot-Savart law, telling us that the magnetic field B of a steady line current I is given by the following formula:

$$B = \frac{\mu_0}{4\pi} \int \frac{I \times x}{\|x\|^3}$$

By applying the curl operator to this formula, we obtain:

$$\begin{aligned}
 \nabla \times B &= \frac{\mu_0}{4\pi} \int \nabla \times \frac{I \times x}{||x||^3} \\
 &= \frac{\mu_0}{4\pi} \int \left\langle \nabla, \frac{x}{||x||^3} \right\rangle J - \langle \nabla, J \rangle \frac{x}{||x||^3} \\
 &= \frac{\mu_0}{4\pi} \int 4\pi \delta_x \cdot J - \frac{\mu_0}{4\pi} \cdot 0 \\
 &= \mu_0 \int \delta_x \cdot J \\
 &= \mu_0 J
 \end{aligned}$$

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

As a conclusion to all this, the equations of magnetostatics are as follows:

THEOREM 3.15. *The equations of magnetostatics are*

$$\langle \nabla, B \rangle = 0$$

$$\nabla \times B = \mu_0 J$$

with the second equation being the Ampère law.

PROOF. This follows indeed from the above discussion, and more specifically from Theorem 3.13 and Theorem 3.14, which both follow from the Biot-Savart law. \square

Observe the obvious analogy with the Gauss equations of electrostatics, namely:

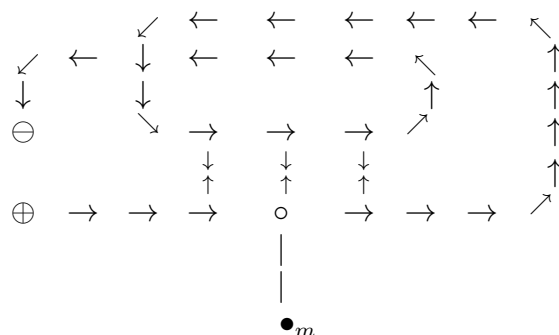
$$\langle \nabla, E \rangle = \frac{\rho}{\varepsilon_0}$$

$$\nabla \times E = 0$$

As a conclusion to all this, looks like someone has played here with basic 3D math, vectors, products and so on, and messed them up, as for electrostatics to become magnetostatics, and vice versa. More on this later, when talking about unification.

As an application of all this, let us discuss motors. The simplest idea of an electromagnetic motor comes from the attracting cables in Fact 3.1, but unfortunately, when doing the engineering and building such a motor, we are led to:

THEOREM 3.16. *A basic electromagnetic motor, pulling a weight m by using the attraction of two parallel currents, travelling in the same direction,*



will not work.

PROOF. Want it or not, this comes from the math in Theorem 3.5, telling us that magnetic forces do not work. And you can't beat such simple math. We will see however, later, that functioning and reliable electromagnetic motors can be built, by tricking the 3D in Theorem 3.5, the idea being that of replacing straight wires by coils. \square

Getting back now to work, and to serious engineering, for being complete, and also for keeping our promises, made at the beginning of this chapter, it remains to discuss magnets. But let us do this slowly. In view of the clear analogy between electrostatics and magnetostatics, it makes sense to investigate first the problem for electrostatics, and then for magnetostatics. And here, we first have, in the context of electrostatics:

FACT 3.17. *When exposed to small electric fields, non-conductors are subject to polarization, coming from small electric rearrangements inside their atoms or molecules.*

To be more precise here, in order to have some discussion started, let us recall from our study of the Faraday cage, from chapter 2, that materials basically fall into two classes, insulators and conductors, and with the conductors being subject to a surface phenomenon of induced charge, when exposed to an electric field.

In the case of insulators, or non-conductors in general, this surface phenomenon of induced charge does not appear, since electrons are strongly bound to their respective atoms or molecules. Of course, when the electric field is very strong, it can blow away electrons, and the material gets ionized. But for lower intensity fields, in the lack of ionization or of an induced surface charge phenomenon, the only thing that can happen is some small rearrangement inside the atoms or molecules, known as polarization.

At the atomic level, the polarization can be understood as being a modification of the electron cloud, or rather of the density of the electron cloud, in the direction of the field.

Here is for illustration a picture of polarized boron ${}_5\text{B}$, exaggerated of course:



For molecules things are far more complicated, because due to the complex geometry of such molecules, the polarization can appear in various 3D directions with respect to the orientation of the field, including the direction orthogonal to it.

This was for the basics of polarization. For more, we refer as usual to any of our standard undergraduate books, such as Griffiths [44]. In what concerns us we will be back to this only later on, when talking about polarized light.

Moving ahead now with magnetism, we have here:

FACT 3.18. When exposed to small magnetic fields, materials are subject to magnetization, similarly to polarization. However, certain special materials like iron ${}_{26}\text{Fe}$ can conserve their magnetization, even after the original magnetizing field is gone.

In what regards the basics of magnetization, these are a bit similar to the basics of polarization, with some physics and geometry going on there, and with the magnetized materials being roughly classified into paramagnets, whose magnetization appears along the field, and diamagnets, whose magnetization appears against the field.

Regarding however the permanent magnets, those which keep their magnetization after the magnetizing field is gone, and which are called ferromagnets, the story here is far more complicated. The general idea is that such materials, including the iron ${}_{26}\text{Fe}$, have the remarkable property of being permanently damaged by a magnetic field, due to a number of quite complicated reasons. In addition, temperature has something to do with all this too, with high temperatures, above the so-called Curie point of the material, which is $T = 770^\circ \text{C}$ for iron, turning ferromagnets into paramagnets.

Again, for more on all this, we refer here as usual to any of our standard books, such as Griffiths [44]. More on metals can be learned by doing some survival-type metallurgy, using the Curie point when cooking, or of course by listening to Metallica.

As a conclusion, we have now a full theory of magnetostatics, complementing the theory of electrostatics. We will see in a moment that these theories can be unified.

3c. Faraday, Maxwell

In the context of moving charges, some of the laws that we know well from electrostatics and from magnetostatics must be altered. But let us first begin with the basics, by forgetting the ideal void that we are used to, and which will be back in a moment, no worries for that. A first question is that of understanding the current density J flowing through a given material, and the answer here is given by Ohm's law, as follows:

FACT 3.19 (Ohm's law). *The current density J is given by*

$$J = \sigma E$$

where σ is a constant, called conductivity of the material.

We are already a bit familiar with this, with our notion of ideal conductor corresponding to $\sigma = \infty$, and our notion of ideal insulator corresponding to $\sigma = 0$. In real life, however, we have of course $\sigma \in (0, \infty)$. Here are 3 + 3 + 3 basic examples, at 20° C and 1 atm, consisting of 3 conductors, 3 semiconductors and 3 insulators, and with σ being replaced by its inverse $\rho = 1/\sigma$, called resistivity, more employed in engineering:

Silver	:	1.59×10^{-8}
Iron	:	9.61×10^{-8}
Graphite	:	1.6×10^{-5}
—		
Seawater	:	0.2
Diamond	:	2.7
Silicon	:	2500
—		
Water	:	8300
Glass	:	$10^9 - 10^{14}$
Teflon	:	$10^{22} - 10^{24}$

Getting back now to Ohm's law, a more familiar version of it is as follows, expressing the total current flowing from one electrode to the other in terms of the potential difference between them, or rather vice versa, and with $R \sim \rho$ being the resistance, which depends, besides on ρ , on the precise configuration of the resistor to be crossed:

$$V = IR$$

With this second formulation of the Ohm law in hand, we can now formulate as well, following Joule, a formula in regards with energy, as follows:

FACT 3.20 (Joule heating law). *The work done by the electric force is*

$$P = VI = I^2R$$

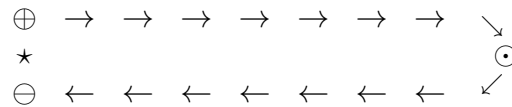
with this being understood as corresponding to heating the resistor.

Let us go back now to the void, with the aim of suitably fixing and unifying the equations of electrostatics and magnetostatics that we have, in the dynamic setting. In fact, there is no need of going through resistors and heating, because we have:

FACT 3.21 (Faraday laws). *The following happen:*

- (1) *Moving a wire loop γ through a magnetic field B produces a current through γ .*
- (2) *Keeping γ fixed, but changing the strength of B , produces too current through γ .*

In order to understand what is going on here, let us start with the simplest electric loop that we know, namely a battery feeding a light bulb:



Here the star stands for the fact that we don't really know what happens inside the battery, typically a complicated chemical process. Nor we will actually worry about the bulb, let us simply assume that this bulb does not exist at all. We will be interested in the force driving the current around the loop, and we have here:

PROPOSITION 3.22. *When writing the force driving the current through a loop γ as*

$$F = F_{\star} + F_e$$

with F_{\star} coming from the source, and F_e coming from the loop, the quantity

$$\mathcal{E} = \int_{\gamma} \langle F(x), dx \rangle$$

called electromotive force, or emf of the loop, is simply obtained by integrating F_{\star} .

PROOF. We have indeed the following computation, based on the fact that F_e being an electrostatic force, its integral over the loop vanishes:

$$\begin{aligned} \mathcal{E} &= \int_{\gamma} \langle F(x), dx \rangle \\ &= \int_{\gamma} \langle F_{\star}(x), dx \rangle + \int_{\gamma} \langle F_e(x), dx \rangle \\ &= \int_{\gamma} \langle F_{\star}(x), dx \rangle + 0 \\ &= \int_{\gamma} \langle F_{\star}(x), dx \rangle \end{aligned}$$

Thus, we have our result, and with the remark of course that the emf $\mathcal{E} \in \mathbb{R}$ is not really a force, but this is the standard terminology, and we will use it. \square

In relation now with the Faraday principles from Fact 3.21, these can be fine-tuned, and reformulated in terms of the emf, in the following way:

FACT 3.23 (Faraday). *The emf of a loop γ moving through a magnetic field B is*

$$\mathcal{E} = -\dot{\Phi}$$

where Φ is the flux of the field B through the loop γ , given by:

$$\Phi = \int_{\gamma} \langle B(x), dx \rangle$$

As for the emf of a fixed loop γ in a changing magnetic field B , this is

$$\mathcal{E} = - \int_{\gamma} \langle \dot{B}(x), dx \rangle$$

which by Stokes is equivalent to the Faraday law $\Delta \times E = -\dot{B}$.

We have now all needed tools for constructing generators and motors, and working out their numerics. Forgetting about generators, the best method for producing energy being $E = mc^2$ anyway, we can now build functioning electric motors, as follows:

THEOREM 3.24. *Functioning and reliable electric motors can be build by using the basic principles of electromagnetism, by using coils of wire.*

PROOF. This is something which improves our previous attempt of building such a motor, reported in Theorem 3.16. To be more precise, that attempt failed due to the math in Theorem 3.5, telling us that magnetic forces do not work. But now we know from Fact 3.23 how to trick Theorem 3.5, by replacing straight wires by loops, or even better, by coils of wire. As for the math and numerics of these motors, these can be worked out too, once again by using the Faraday formulae from Fact 3.23. \square

There are of course tons of other things that can be said about electromechanics, and we refer here to any of our standard undergraduate books on electrodynamics, such as Griffiths [44]. And with the remark however that, for serious applications, you are in need afterwards of a solid engineering book, centered on electromechanics.

Getting back to theory, the above considerations lead to the following conclusion:

FACT 3.25 (Faraday). *In the context of moving chages, the electrostatics law*

$$\nabla \times E = 0$$

must be replaced by the following equation,

$$\nabla \times E = -\dot{B}$$

called Faraday law.

Along the same lines, and following now Maxwell, there is a correction as well to be made to the main law of magnetostatics, namely the Ampère law, as follows:

FACT 3.26 (Maxwell). *In the context of moving chages, the Ampère law*

$$\nabla \times B = \mu_0 J$$

must be replaced by the following equation,

$$\nabla \times B = \mu_0(J + \varepsilon_0 \dot{E})$$

called Ampère law with Maxwell correction term.

Now by putting everything together, and perhaps after doublechecking as well, with all sorts of experiments, that the remaining electrostatics and magnetostatics laws, that we have not modified, work indeed fine in the dynamic setting, we obtain:

THEOREM 3.27 (Maxwell). *Electrodynamics is governed by the formulae*

$$\langle \nabla, E \rangle = \frac{\rho}{\varepsilon_0}$$

$$\langle \nabla, B \rangle = 0$$

$$\nabla \times E = -\dot{B}$$

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

called Maxwell equations.

PROOF. This follows indeed from the above, the details being as follows:

- (1) The first equation is the Gauss law, that we know well from chapter 1.
- (2) The second equation is something anonymous, that we know well too.
- (3) The third equation is a previously anonymous law, modified into Faraday's law.
- (4) And the fourth equation is the Ampère law, as modified by Maxwell. □

3d. Lorentz invariance

At the level of general theory now, there are many things that can be said, as a continuation of the above. In what concerns us, we will be mostly theoretical. As a first key result, making the connection with Einstein's relativity theory, we have:

THEOREM 3.28. *The Maxwell equations are invariant under Lorentz transformations*

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

$$y' = y$$

$$z' = z$$

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

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

PROOF. This is something a bit complicated, the idea being as follows:

(1) As a first comment, this result, due to Lorentz himself, working on electromagnetism, was established some time before Einstein's relativity theory.

(2) As for the 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:

$$\begin{aligned} E'_x &= E_x \\ E'_y &= \gamma(E_y - vB_z) \\ E'_z &= \gamma(E_z + vB_y) \end{aligned}$$

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

$$\begin{aligned} B'_x &= B_x \\ B'_y &= \gamma\left(B_y + \frac{v}{c^2}E_z\right) \\ B'_z &= \gamma\left(B_z - \frac{v}{c^2}E_y\right) \end{aligned}$$

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

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

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

$$\begin{aligned} E' &= DE + cMB \\ B' &= DB - \frac{M}{c}E \end{aligned}$$

(4) 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. \square

There are many other things that can be said here, in relation with Einstein's relativity, which are all very interesting. In fact, the discussion here is quite endless. We will come back to this later in this book, on numerous occasions.

3e. Exercises

Exercises.

CHAPTER 4

Radiation and waves

4a. Wave equation

The idea is that light is an electromagnetic wave. Thus, we first need to talk about waves. The result that we will need, from classical mechanics, is as follows:

THEOREM 4.1. *The wave equation is*

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

where $\Delta = \sum_i d^2/dx_i^2$ is the Laplace operator.

PROOF. There are several proofs here, a nice one, by discretizing, being 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:

$$\dots \times \times \times \bullet \times \times \times \bullet \times \times \times \bullet \times \times \times \bullet \times \times \times \bullet \times \times \times \bullet \times \times \times \dots$$

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

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

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 $N \gg 0$ 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 $N \rightarrow \infty$, and therefore $l \rightarrow 0$, we obtain in this way:

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

(3) In arbitrary N dimensions now, the same argument carries on, and we are led to the following equation, with $v = \sqrt{K/M} \cdot L$ being the propagation speed:

$$\ddot{\varphi}(x) = v^2 \sum_i \frac{d^2\varphi}{dx_i^2}(x)$$

But we recognize at right the Laplace operator, and we are done. There is of course some more discussion to be made here, arguing that our spring model in (1) is indeed the correct one, for modelling such wave propagation questions. But hey, we're doing theoretical physics here. And don't worry, experiments confirm our findings. \square

Going now towards light, we will need as well a second piece of physics, as follows:

THEOREM 4.2 (Maxwell theory). *In regions of space where there is no charge or current present the Maxwell equations for electrodynamics read*

$$\langle \nabla, E \rangle = \langle \nabla, B \rangle = 0$$

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

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 $c = 299,792,458$.

PROOF. We can see right away that under the circumstances in the statement, the Maxwell equations take a simple and comprehensible form, as follows:

$$\langle \nabla, E \rangle = \langle \nabla, B \rangle = 0$$

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

On the other hand, we know from Biot-Savart that the constant involved is:

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

Summarizing, we have our equations. Leaving aside the first two equations, by applying the curl operator to the last two equations, we obtain:

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

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

$$\nabla \times (\nabla \times \varphi) = \nabla \langle \nabla, \varphi \rangle - \Delta \varphi$$

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

So, what is light? Light is the wave predicted by Theorem 4.2, travelling at 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.

4b. Radiation, light

In what regards the creation of light, the mechanism here is as follows:

FACT 4.3. *An accelerating or decelerating charge produces electromagnetic radiation, called light, whose frequency and wavelength can be explicitly computed.*

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

In view of the above, and especially of the light bulb example, a natural question appears, that you surely have on your mind, right now: what about a resistor which is not a light bulb filament, what happens to the light produced there?

This is a good question, and we already know a part of answer to it, coming from the Joule law, which says that the resistor will start heating. The second part of it, that we will discuss in a moment, states that responsible for heat is, guess who, light again, but this time in non-visible wavelengths, typically IR.

Getting back now to Fact 4.3, in general, as stated above, this is something which can be deduced via some math, based on the Maxwell equations. However, all this math is not exactly trivial, and for details here, you can see for instance Griffiths [44].

Let us go back now to the wave equation $\ddot{\varphi} = v^2 \Delta \varphi$ from Theorem 4.2, and try to understand its simplest solutions. In 1D, the situation is as follows:

THEOREM 4.4. *The 1D wave equation, with speed v , namely*

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

has as basic solutions the following functions,

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

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

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

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

PROOF. 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.

(3) Finally, the last assertion is something standard, coming from Fourier analysis, that we will not really need, in what follows. \square

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:

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

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:

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

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

DEFINITION 4.5. *A monochromatic plane wave is a solution of the 3D wave equation which moves in only 1 direction, making it in practice a solution of the 1D wave equation, and which is of the special form found in Theorem 4.4, with no frequencies mixed.*

In other words, we are making here two assumptions on our wave. First is the 1-dimensionality assumption, which gets us into the framework of Theorem 4.4. And second is the assumption, in connection with the Fourier decomposition result from the end of Theorem 4.4, that our solution is of “pure” type, meaning a wave having a well-defined wavelength 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 4.2 and Fact 4.3, we have:

FACT 4.6. *Physically speaking, a monochromatic plane wave is the electromagnetic radiation appearing as in Theorem 4.2 and Fact 4.3, via equations of type*

$$E = \operatorname{Re}(\tilde{E}) \quad : \quad \tilde{E} = \tilde{E}_0 e^{i(\langle k, x \rangle - wt)}$$

$$B = \operatorname{Re}(\tilde{B}) \quad : \quad \tilde{B} = \tilde{B}_0 e^{i(\langle k, x \rangle - wt)}$$

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 4.4 and Definition 4.5, into the context of our original physics discussion, from Theorem 4.2 and Fact 4.3. And also to add an extra twist coming from physics, and more specifically from the notion of polarization.

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

the familiar term “light”. The classification, in a rough form, is as follows:

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

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

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

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

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

Importantly, both ends of the table are a bit loose. At the high frequency end there are some restrictions coming from quantum mechanics, and more on them later. As for the low frequency end, what’s wave and what’s not is a bit of a philosophical question, but which is actually not that philosophical, because waves having huge wavelengths can

easily turn around mountains, full countries and so on, and so are of military interest. Secret research here, more of engineering type of course, is still ongoing.

Back now to our business, with all the above in hand, we can do some optics. Light usually comes in “bundles”, with waves of several wavelengths coming at the same time, from the same source, and the first challenge is that of separating these wavelengths. In order to discuss this, let us start with the following fact:

FACT 4.7. *Inside a linear, homogeneous medium, where there is no free charge or current present, the Maxwell equations for electrodynamics read*

$$\langle \nabla, E \rangle = \langle \nabla, B \rangle = 0$$

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

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

Observe that this is precisely the first part of Theorem 4.2, with the vacuum constants ε_0, μ_0 being replaced by their versions ε, μ , concerning the medium in question. In what regards now the second part of Theorem 4.2, which was a theorem, we have:

THEOREM 4.8. *Inside a linear, homogeneous medium, where there is no free charge or free current present, both E and B are subject to the wave equation*

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

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

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

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

PROOF. This is something that we know well in vacuum, and the proof in general is identical, with the resulting speed being:

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

But this formula can be written in a more familiar form, as above. □

As a first observation here, while the above is something quite trivial, mathematically speaking, from the physical viewpoint we are here into complicated things. Materials can be transparent or opaque, with the distinction between them being something very subtle, and advanced, and Theorem 4.8 obviously deals with the transparent case.

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

FACT 4.9. *When travelling 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 [44].

As a simple consequence of the above, we have:

THEOREM 4.10. *Light can be decomposed, by using a prism.*

PROOF. This follows from Fact 4.9. Indeed, when hitting a piece of glass, provided that the hitting angle is not 90° , the light will decompose over the wavelengths 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. \square

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

FACT 4.11. *We can study events via spectroscopy, by capturing the light the event has produced, decomposing it with a prism, carefully recording its “spectral signature”, consisting of the wavelengths present, and their density, and then doing some reverse engineering, consisting in reconstructing the event out of its spectral signature.*

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

4c. Spectral lines

As an application of the above, let us discuss the spectral lines of hydrogen H. There is a long story here, involving many discoveries of many people, around 1890-1900. First on our list is the following discovery, which actually came second, by Lyman in 1906:

FACT 4.12 (Lyman). *The hydrogen atom has spectral lines given by the formula*

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

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

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

called *Lyman series of the hydrogen atom*.

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

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

FACT 4.13 (Balmer). *The hydrogen atom has spectral lines given by the formula*

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

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

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

called *Balmer series of the hydrogen atom*.

So, this was Balmer's original result, which started everything, and with his original wavelength formula being in fact something equivalent to the above formula, but a bit

more complicated, as follows, with $B \simeq 3.645 \times 10^{-7}$ being the Balmer constant:

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

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

FACT 4.14 (Paschen). *The hydrogen atom has spectral lines given by the formula*

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

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

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

called *Paschen series of the hydrogen atom*.

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

CONCLUSION 4.15 (Rydberg, Ritz). *The spectral lines of the hydrogen atom are given by the Rydberg formula, depending on integer parameters $n_1 \leq n_2$,*

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

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:

$$\frac{1}{\lambda_{n_1 n_2}} + \frac{1}{\lambda_{n_2 n_3}} = \frac{1}{\lambda_{n_1 n_3}}$$

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 $n_1 = 4, 5, 6, \dots$, and these were discovered

later, by Brackett in 1922, Pfund in 1924, Humphreys in 1953, and others afterwards, 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 – ∞	Lyman	91.13 nm	UV
2	3 – ∞	Balmer	364.51 nm	UV
3	4 – ∞	Paschen	820.14 nm	IR
		—	—	
4	5 – ∞	Brackett	1458.03 nm	far IR
5	6 – ∞	Pfund	2278.17 nm	far IR
6	7 – ∞	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 $R \rightarrow RZ^2$, where Z is the atomic number.

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

THOUGHT 4.16. *The simplest observables of the hydrogen atom, combining via*

$$\frac{1}{\lambda_{n_1 n_2}} + \frac{1}{\lambda_{n_2 n_3}} = \frac{1}{\lambda_{n_1 n_3}}$$

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

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

THOUGHT 4.17. *The Ritz-Rydberg combination principle reminds the formula*

$$e_{n_1 n_2} e_{n_2 n_3} = e_{n_1 n_3}$$

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

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

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

This looks certainly very interesting, and actually reminds relativity. Remember the Galileo addition formula there for relative speeds, which gets converted into the Einstein

addition formula for the same relative speeds? Well, the reasoning there was as follows, involving the same sort of weird additions as above:

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

In short, we are in familiar territory here, and we can start dreaming of:

THOUGHT 4.18. *Observables in quantum mechanics should be some sort of infinite matrices, generalizing the Lyman, Balmer, Paschen lines of the hydrogen atom, and multiplying between them as the matrices do, as to produce further observables.*

And probably enough for now, this is the kind of discovery that should be celebrated with slaughtering 50 sheep and inviting your friends over, for a banquet, as the legend goes that Pythagoras did, after he discovered his $a^2 + b^2 = c^2$ theorem.

Summarizing, we have seen that many interesting things can be said about matter, via spectroscopy and light, and with all these experiments and data still in need to be processed, via a clever theory of “quantum mechanics”, which remains to be invented.

Time now to put everything together. 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, let us agree on:

CLAIM 4.19. *Ordinary matter is made of small particles called atoms, with each atom appearing as a mix of even smaller particles, namely protons +, neutrons 0 and electrons -, with the same number of protons + and electrons -.*

As a first observation, this is something which does not look obvious at all, with probably lots of work, by many people, being involved, as to lead to this claim. And so it is. The story goes back to the discovery of charges and electricity, which were attributed to a small particle, the electron -. Now since matter is by default neutral, this naturally leads to the consideration to the proton +, having the same charge as the electron.

But, 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 4.19.

Of particular importance here was as well the radioactivity theory of Becquerel and Pierre and Marie Curie, involving this time such small beasts, or perhaps some related radiation, peeling off by themselves, in heavy elements such as uranium ${}_{92}\text{U}$, polonium ${}_{84}\text{Po}$ and radium ${}_{88}\text{Ra}$. 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 4.19, with further particles involved, and more on this later, but as a general idea, all this deluge of small particle findings, all coming around 1900-1910, further solidified Claim 4.19.

So, taking now Claim 4.19 for granted, how are then the atoms organized, as mixtures of protons +, neutrons 0 and electrons –? The answer here lies again in the above-mentioned “brutal” experiments of Thomson, Rutherford and others, which not only proved Claim 4.19, but led to an improved version of it, as follows:

CLAIM 4.20. *The atoms are formed by a core of protons + and neutrons 0, surrounded by a cloud of electrons –, gravitating around the core.*

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

CLAIM 4.21 (Bohr and others). *The atoms are formed by a core of protons and neutrons, surrounded by a cloud of electrons, basically obeying to a modified version of electromagnetism. And with a fine mechanism involved, as follows:*

- (1) *The electrons are free to move only on certain specified elliptic orbits, labelled $1, 2, 3, \dots$, 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 $n_1 \rightarrow n_2$ or $n_2 \rightarrow n_1$, is given, via the Planck viewpoint, by the Rydberg formula, applied with $n_1 < n_2$.*
- (4) *The simplest such jumps are those observed by Lyman, Balmer, Paschen. And multiple jumps explain the Ritz-Rydberg formula.*

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

And the story is not over here. Following now Heisenberg, the next claim is that the underlying math in all the above can lead to a beautiful axiomatization of quantum mechanics, as a “matrix mechanics”, along the lines of Thought 4.18.

In practice now, quantum mechanics can be developed along these lines, and this was something which was done by Heisenberg and Schrödinger, using various methods, and then Dirac and others. Among their main discoveries, obtained by solving the hydrogen atom along the lines of Claim 4.21, is the fact that the electron orbits depend not only on the energy level $n \in \mathbb{N}$, as above, but in fact on a triple of parameters n, l, m , with the extra parameters l, m having something to do with the geometry of the orbit:

$$n \rightarrow (n, l, m)$$

On top of this, comes the spin of the electron, which is something that we ignored so far. And spin is in fact a key component to our problem, because we have:

FACT 4.22 (Pauli exclusion principle). *Two electrons cannot occupy the same quantum numbers n, l, m , with same spin $s = \pm 1/2$.*

But this Pauli exclusion principle gives in fact the key to the understanding of the arbitrary, $Z \geq 2$ 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, $Z = 1, \dots, 118$. 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	$\frac{\text{H}}{1}$																		$\frac{\text{He}}{2}$
2	$\frac{\text{Li}}{3}$	$\frac{\text{Be}}{4}$												$\frac{\text{B}}{5}$	$\frac{\text{C}}{6}$	$\frac{\text{N}}{7}$	$\frac{\text{O}}{8}$	$\frac{\text{F}}{9}$	$\frac{\text{Ne}}{10}$
3	$\frac{\text{Na}}{11}$	$\frac{\text{Mg}}{12}$												$\frac{\text{Al}}{13}$	$\frac{\text{Si}}{14}$	$\frac{\text{P}}{15}$	$\frac{\text{S}}{16}$	$\frac{\text{Cl}}{17}$	$\frac{\text{Ar}}{18}$
4	$\frac{\text{K}}{19}$	$\frac{\text{Ca}}{20}$		$\frac{\text{Sc}}{21}$	$\frac{\text{Ti}}{22}$	$\frac{\text{V}}{23}$	$\frac{\text{Cr}}{24}$	$\frac{\text{Mn}}{25}$	$\frac{\text{Fe}}{26}$	$\frac{\text{Co}}{27}$	$\frac{\text{Ni}}{28}$	$\frac{\text{Cu}}{29}$	$\frac{\text{Zn}}{30}$	$\frac{\text{Ga}}{31}$	$\frac{\text{Ge}}{32}$	$\frac{\text{As}}{33}$	$\frac{\text{Se}}{34}$	$\frac{\text{Br}}{35}$	$\frac{\text{Kr}}{36}$
5	$\frac{\text{Rb}}{37}$	$\frac{\text{Sr}}{38}$		$\frac{\text{Y}}{39}$	$\frac{\text{Zr}}{40}$	$\frac{\text{Nb}}{41}$	$\frac{\text{Mo}}{42}$	$\frac{\text{Tc}}{43}$	$\frac{\text{Ru}}{44}$	$\frac{\text{Rh}}{45}$	$\frac{\text{Pd}}{46}$	$\frac{\text{Ag}}{47}$	$\frac{\text{Cd}}{48}$	$\frac{\text{In}}{49}$	$\frac{\text{Sn}}{50}$	$\frac{\text{Sb}}{51}$	$\frac{\text{Te}}{52}$	$\frac{\text{I}}{53}$	$\frac{\text{Xe}}{54}$
6	$\frac{\text{Cs}}{55}$	$\frac{\text{Ba}}{56}$	l	$\frac{\text{Lu}}{71}$	$\frac{\text{Hf}}{72}$	$\frac{\text{Ta}}{73}$	$\frac{\text{W}}{74}$	$\frac{\text{Re}}{75}$	$\frac{\text{Os}}{76}$	$\frac{\text{Ir}}{77}$	$\frac{\text{Pt}}{78}$	$\frac{\text{Au}}{79}$	$\frac{\text{Hg}}{80}$	$\frac{\text{Tl}}{81}$	$\frac{\text{Pb}}{82}$	$\frac{\text{Bi}}{83}$	$\frac{\text{Po}}{84}$	$\frac{\text{At}}{85}$	$\frac{\text{Rn}}{86}$
7	$\frac{\text{Fr}}{87}$	$\frac{\text{Ra}}{88}$	a	$\frac{\text{Lr}}{103}$	$\frac{\text{Rf}}{104}$	$\frac{\text{Db}}{105}$	$\frac{\text{Sg}}{106}$	$\frac{\text{Bh}}{107}$	$\frac{\text{Hs}}{108}$	$\frac{\text{Mt}}{109}$	$\frac{\text{Ds}}{110}$	$\frac{\text{Rg}}{111}$	$\frac{\text{Cn}}{112}$	$\frac{\text{Nh}}{113}$	$\frac{\text{Fl}}{114}$	$\frac{\text{Mc}}{115}$	$\frac{\text{Lv}}{116}$	$\frac{\text{Ts}}{117}$	$\frac{\text{Og}}{118}$
			l :	$\frac{\text{La}}{57}$	$\frac{\text{Ce}}{58}$	$\frac{\text{Pr}}{59}$	$\frac{\text{Nd}}{60}$	$\frac{\text{Pm}}{61}$	$\frac{\text{Sm}}{62}$	$\frac{\text{Eu}}{63}$	$\frac{\text{Gd}}{64}$	$\frac{\text{Tb}}{65}$	$\frac{\text{Dy}}{66}$	$\frac{\text{Ho}}{67}$	$\frac{\text{Er}}{68}$	$\frac{\text{Tm}}{69}$	$\frac{\text{Yb}}{70}$		
			a :	$\frac{\text{Ac}}{89}$	$\frac{\text{Th}}{90}$	$\frac{\text{Pa}}{91}$	$\frac{\text{U}}{92}$	$\frac{\text{Np}}{93}$	$\frac{\text{Pu}}{94}$	$\frac{\text{Am}}{95}$	$\frac{\text{Cm}}{96}$	$\frac{\text{Bk}}{97}$	$\frac{\text{Cf}}{98}$	$\frac{\text{Es}}{99}$	$\frac{\text{Fm}}{100}$	$\frac{\text{Md}}{101}$	$\frac{\text{No}}{102}$		

Here the horizontal parameter $1, \dots, 18$ is called the group, and the vertical parameter $1, \dots, 7$ is called the period. The two rows on the bottom consist of lanthanum $_{57}\text{La}$ and its followers, called lanthanides, and of actinium $_{89}\text{Ac}$ and its followers, called actinides. These are to be inserted in the main table, where indicated, lanthanides between barium $_{56}\text{Ba}$ and lutetium $_{71}\text{Lu}$, and actinides between radium $_{88}\text{Ra}$ and lawrencium $_{103}\text{Lr}$.

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 $1, \dots, 18$. Their group is by definition "lanthanides" and "actinides".

We will comment in a moment on all this, but before anything:

ADVICE 4.23. *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 ${}_{36}\text{Kr}$, which are absolutely needed for everything, and must be all learned, to start with:

NAMES 4.24. *The elements up to krypton ${}_{36}\text{Kr}$ are as follows:*

- (1) *Hydrogen ${}_{1}\text{H}$, helium ${}_{2}\text{He}$.*
- (2) *Lithium ${}_{3}\text{Li}$, beryllium ${}_{4}\text{Be}$, boron ${}_{5}\text{B}$, carbon ${}_{6}\text{C}$, nitrogen ${}_{7}\text{N}$, oxygen ${}_{8}\text{O}$, fluorine ${}_{9}\text{F}$, neon ${}_{10}\text{Ne}$.*
- (3) *Sodium ${}_{11}\text{Na}$, magnesium ${}_{12}\text{Mg}$, aluminium ${}_{13}\text{Al}$, silicon ${}_{14}\text{Si}$, phosphorus ${}_{15}\text{P}$, sulfur ${}_{16}\text{S}$, chlorine ${}_{17}\text{Cl}$, argon ${}_{18}\text{Ar}$.*
- (4) *Potassium ${}_{19}\text{K}$, calcium ${}_{20}\text{Ca}$, scandium ${}_{21}\text{Sc}$, titanium ${}_{22}\text{Ti}$, vanadium ${}_{23}\text{V}$, chromium ${}_{24}\text{Cr}$, manganese ${}_{25}\text{Mn}$, iron ${}_{26}\text{Fe}$, cobalt ${}_{27}\text{Co}$.*
- (5) *Nickel ${}_{28}\text{Ni}$, copper ${}_{29}\text{Cu}$, zinc ${}_{30}\text{Zn}$, gallium ${}_{31}\text{Ga}$, germanium ${}_{32}\text{Ge}$, arsenic ${}_{33}\text{As}$, selenium ${}_{34}\text{Se}$, bromine ${}_{35}\text{Br}$, krypton ${}_{36}\text{Kr}$.*

Observe that all names fit with the abbreviations, except for sodium ${}_{11}\text{Na}$, coming from the Latin natrium, potassium ${}_{19}\text{K}$, coming from the Latin kalium, iron ${}_{26}\text{Fe}$ coming from the Latin ferrum, and also copper ${}_{29}\text{Cu}$, coming from the Latin cuprum.

In what regards the elements heavier than krypton ${}_{36}\text{Kr}$, it is heartbreaking to sort them out, but as a useful complement to the above list, we have:

NAMES 4.25. *Remarkable elements heavier than krypton ${}_{36}\text{Kr}$ include:*

- (1) *Noble gases: xenon ${}_{54}\text{Xe}$, radon ${}_{86}\text{Rn}$.*
- (2) *Noble metals: silver ${}_{47}\text{Ag}$, iridium ${}_{77}\text{Ir}$, platinum ${}_{78}\text{Pt}$, gold ${}_{47}\text{Au}$.*
- (3) *Heavy metals: mercury ${}_{80}\text{Hg}$, lead ${}_{82}\text{Pb}$.*
- (4) *Radioactive: polonium ${}_{84}\text{Po}$, radium ${}_{88}\text{Ra}$, uranium ${}_{92}\text{U}$, plutonium ${}_{94}\text{Pu}$.*
- (5) *Miscellaneous: rubidium ${}_{37}\text{Rb}$, strontium ${}_{38}\text{Sr}$, molybdenum ${}_{42}\text{Mo}$, technetium ${}_{43}\text{Tc}$, cadmium ${}_{48}\text{Cd}$, tin ${}_{50}\text{Sn}$, iodine ${}_{53}\text{I}$, caesium ${}_{55}\text{Cs}$, tungsten ${}_{74}\text{Tu}$, bismuth ${}_{83}\text{Bi}$, francium ${}_{87}\text{Fr}$, americium ${}_{95}\text{Am}$.*

Here the abbreviations not fitting with English names come from the Latin or sometimes Greek argentum ${}_{47}\text{Ag}$, aurum ${}_{47}\text{Au}$, hydrargyrum ${}_{80}\text{Hg}$, plumbum ${}_{82}\text{Pb}$ and stannum ${}_{50}\text{Sn}$. The noble gases in (1) normally include oganesson ${}_{118}\text{Og}$ 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 ${}_{43}\text{Tc}$ 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 4.26. *Any $Z = 1, \dots, 118$ 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 these versions are “family”, and the element 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 4.27. *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 $1, \dots, 7$ corresponds to the highest n occupied, and the group $1, \dots, 18$ comes from l, m, s .*

This is of course something very basic, and there is a detailed analysis to be done afterwards, for $Z = 1, \dots, 118$. For the elements up to krypton ${}_{36}\text{Kr}$, the list of electron configurations can be found for instance in Feynman [39] or Griffiths [45].

With this in hand, we can now start doing some chemistry. There is an enormous quantity of things that can be said here, the simplest of which being:

THEOREM 4.28. *The group 18 elements, helium ${}_{2}\text{He}$, neon ${}_{10}\text{Ne}$, argon ${}_{18}\text{Ar}$, krypton ${}_{36}\text{Kr}$, xenon ${}_{54}\text{Xe}$ and radon ${}_{86}\text{Rn}$, called noble gases, are allergic to chemistry.*

PROOF. This follows from the above discussion, 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 ${}_{118}\text{Og}$ 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

So long for the chemical elements, and the periodic table. Unfortunately business is business, and we will have to stop here, and go back to electrodynamics.

4d. Max Planck

Our discussion so far implicitly used the fact that heat is light too, and so time now, before getting into quantum mechanics, to get back to the theory of heat, as developed before, and see what our new viewpoint on it can bring. The main problem here is to compute the radiation of black bodies, and we will discuss this now.

Consider a black body, that is to say, a body at thermal equilibrium, assumed to be at temperature T . This body radiates heat, and we are interested in computing the energy density of the radiation $\mathcal{E}(\nu, T)$, around a given frequency ν of this radiation.

Quite surprisingly, the intuitive and honest modelling of the problem, and the subsequent math, done honestly too, lead to a spectacularly wrong result, as follows:

THEOREM 4.29. *We have the Rayleigh-Jeans formula for the energy density*

$$\mathcal{E}(\nu, T) = \frac{8\pi bT}{c^3} \nu^2$$

where b is the Boltzmann constant, leading globally to the divergent integral

$$\mathcal{E} = \frac{8\pi bTV}{c^3} \int_0^\infty \nu^2 d\nu$$

over a volume V , with this divergence phenomenon being called *UV catastrophe*.

PROOF. This is arguably the most famous wrong result in the history of physics, so we will spend some time in trying to understand its proof. And with the comment that this will be no waste of time, because the fix, found later by Max Planck, uses exactly the same ideas and computations, but with an unexpected twist at the end.

(1) Our starting point are the equations for the electromagnetic radiation, that we will now regard as heat, as formulated before, namely:

$$E = \text{Re}(\tilde{E}) \quad : \quad \tilde{E} = e_n e^{i(\langle k_n, x \rangle - w_n t)}$$

$$B = \text{Re}(\tilde{B}) \quad : \quad \tilde{B} = b_n e^{i(\langle k_n, x \rangle - w_n t)}$$

Here n is a certain parameter, that will appear later on, and that we can for the moment ignore. Now inserting this data into the Maxwell equations gives the following formulae, connecting the parameters, that we will use several times in what follows:

$$k_n \times b_n + \frac{w_n}{c} e_n = 0$$

$$k_n \times e_n - \frac{w_n}{c} b_n = 0$$

$$\langle k_n, e_n \rangle = \langle k_n, b_n \rangle = 0$$

(2) Let us compute the electromagnetic energy in a finite volume $V = L^3$. We will use here the well-known fact, coming from classical electrodynamics, that the energy density in radiation is $(\|E\|^2 + \|B\|^2)/8\pi$. Thus, the energy we are looking for is given by:

$$\mathcal{E} = \frac{1}{8\pi} \int_V (\|E\|^2 + \|B\|^2)$$

(3) In order to compute this integral, let us better model our question. Due to obvious periodicity reasons, the wave number k and the angular frequency w must be of the following form, with $n \in \mathbb{Z}^3$ being a vector with integer components:

$$k_n = \frac{2\pi}{L} \cdot n \quad , \quad w_n = c\|k_n\|$$

Thus, the electric and magnetic fields in our enclosure $V = L^3$ appear as linear combinations as follows, for certain vectors $e_n, b_n \perp n$, related by the formulae in (1):

$$E = \text{Re}(\tilde{E}) \quad : \quad \tilde{E} = \sum_n e_n e^{i(\langle k_n, x \rangle - w_n t)}$$

$$B = \text{Re}(\tilde{B}) \quad : \quad \tilde{B} = \sum_n b_n e^{i(\langle k_n, x \rangle - w_n t)}$$

(4) According to the above formula of E , we have:

$$\begin{aligned} \|E\|^2 &= \|\text{Re}(\tilde{E})\|^2 \\ &= \frac{1}{4} \left\| \sum_n e_n e^{i(\langle k_n, x \rangle - w_n t)} + \bar{e}_n e^{-i(\langle k_n, x \rangle - w_n t)} \right\|^2 \\ &= \frac{1}{4} \sum_{nm} \langle e_n, e_m \rangle e^{i(\langle k_n - k_m, x \rangle - (w_n - w_m)t)} \\ &\quad + \frac{1}{4} \sum_{nm} \langle e_n, \bar{e}_m \rangle e^{i(\langle k_n + k_m, x \rangle - (w_n + w_m)t)} \\ &\quad + \frac{1}{4} \sum_{nm} \langle \bar{e}_n, e_m \rangle e^{i(-\langle k_n + k_m, x \rangle + (w_n + w_m)t)} \\ &\quad + \frac{1}{4} \sum_{nm} \langle \bar{e}_n, \bar{e}_m \rangle e^{i(-\langle k_n - k_m, x \rangle + (w_n - w_m)t)} \end{aligned}$$

(5) Now by integrating, we obtain the following formula:

$$\begin{aligned} \frac{1}{V} \int_V \|E\|^2 &= \frac{1}{4} \sum_n \langle e_n, e_n \rangle + \frac{1}{4} \sum_n \langle e_n, \bar{e}_{-n} \rangle e^{-2iw_n t} \\ &\quad + \frac{1}{4} \sum_n \langle \bar{e}_n, e_{-n} \rangle e^{2iw_n t} + \frac{1}{4} \sum_n \langle \bar{e}_n, \bar{e}_n \rangle \end{aligned}$$

(6) Similarly, according to the above formula of B , we have:

$$\begin{aligned} \frac{1}{V} \int_V \|B\|^2 &= \frac{1}{4} \sum_n \langle b_n, b_n \rangle + \frac{1}{4} \sum_n \langle b_n, \bar{b}_{-n} \rangle e^{-2iw_n t} \\ &\quad + \frac{1}{4} \sum_n \langle \bar{b}_n, b_{-n} \rangle e^{2iw_n t} + \frac{1}{4} \sum_n \langle \bar{b}_n, \bar{b}_n \rangle \end{aligned}$$

(7) Before summing the integrals that we found, let us use the formulae connecting the parameters k_n, e_n, b_n found in (1) above, namely:

$$k_n \times b_n + \frac{w_n}{c} e_n = 0$$

$$k_n \times e_n - \frac{w_n}{c} b_n = 0$$

$$\langle k_n, e_n \rangle = \langle k_n, b_n \rangle = 0$$

By using these formulae, we obtain the following identities:

$$\langle b_n, b_n \rangle = \frac{c^2}{w_n^2} \langle k_n \times e_n, k_n \times e_n \rangle = \frac{c^2 \|k_n\|^2}{w_n^2} \langle e_n, e_n \rangle = \langle e_n, e_n \rangle$$

$$\langle b_n, \bar{b}_{-n} \rangle = \frac{c^2}{w_n^2} \langle k_n \times e_n, k_{-n} \times \bar{e}_n \rangle = -\frac{c^2 \|k_n\|^2}{w_n^2} \langle e_n, \bar{e}_{-n} \rangle = -\langle e_n, \bar{e}_{-n} \rangle$$

$$\langle \bar{b}_n, b_{-n} \rangle = \frac{c^2}{w_n^2} \langle k_n \times \bar{e}_n, k_{-n} \times e_n \rangle = -\frac{c^2 \|k_n\|^2}{w_n^2} \langle \bar{e}_n, e_{-n} \rangle = -\langle \bar{e}_n, e_{-n} \rangle$$

$$\langle \bar{b}_n, \bar{b}_n \rangle = \frac{c^2}{w_n^2} \langle k_n \times \bar{e}_n, k_n \times \bar{e}_n \rangle = \frac{c^2 \|k_n\|^2}{w_n^2} \langle \bar{e}_n, \bar{e}_n \rangle = \langle \bar{e}_n, \bar{e}_n \rangle$$

(8) We conclude that when summing the integrals computed in (5) and (6), all the terms involving phases will cancel, and we obtain the following formula:

$$\frac{1}{V} \int_V \|E\|^2 + \|B\|^2 = \frac{1}{2} \sum_n \langle e_n, e_n \rangle + \frac{1}{2} \sum_n \langle \bar{e}_n, \bar{e}_n \rangle$$

Now by multiplying everything by $V/8\pi$, as explained in (2), we obtain:

$$\mathcal{E} = \frac{V}{16\pi} \sum_n (\langle e_n, e_n \rangle + \langle \bar{e}_n, \bar{e}_n \rangle)$$

(9) The point now is that, by computing this sum, we are led to the Rayleigh-Jeans formula in the statement for the corresponding radiation energy density, namely:

$$\mathcal{E}(\nu, T) = \frac{8\pi bT}{c^3} \nu^2$$

(10) And this is certainly wrong, because the total energy which is radiated by our black body, all over the frequency spectrum, follows to be:

$$\mathcal{E} = \frac{8\pi bTV}{c^3} \int_0^\infty \nu^2 d\nu = \infty$$

More precisely, the Rayleigh-Jeans formula works quite well all across the frequency spectrum, in particular fitting well with the known data, except for the UV range, where things diverge. And with this phenomenon being called “UV catastrophe”. \square

Well, looks like we are in deep trouble here. Fortunately, the solution to the UV catastrophe, and to the black body problem in general, was found a few years later by Max Plank, his bold new modelling method, and result, being as follows:

THEOREM 4.30. *The correct formula for the black body radiation, obtained by assuming that energy is quantized, is the Planck formula*

$$\mathcal{E}(\nu, T) d\nu = \frac{8\pi h_0}{c^3} \cdot \frac{\nu^3 d\nu}{e^{h_0\nu/bT} - 1}$$

with h_0 being a new constant, called Planck constant. This formula fits with all known data, fits as well with Rayleigh-Jeans outside the UV range, and globally leads to

$$\mathcal{E} = \int_0^\infty \mathcal{E}(\nu, T) d\nu = aT^4$$

with the radiation energy constant on the right being given by:

$$a = \frac{16\pi^8 b^4}{15h_0^3 c^3}$$

PROOF. This is something quite technical, obtained by further building on the formula found in (8) above, but counting this time in a new way, by assuming that the energy is quantized. For details here, and for more, we refer for instance to Weinberg [90]. \square

So, this is the famous Planck formula, whose consequences go far beyond its scope. Indeed, regardless on what the original problem was about, namely black bodies, and who cares after all about them, we have now proof for the fact that the energy is quantized. And this, simply because any attempt of solving the problem without assuming that energy is quantized leads to a catastrophe, as explained before.

A word about the new constant appearing in the above, namely the Planck constant. Due to some technical reasons, we will use here a non-standard notation for this constant, with the figures and our conventions being as follows:

FACT 4.31. *The Planck constant is given by the exact formula*

$$h_0 = 6.626\ 070\ 15 \times 10^{-34}$$

as per latest SI regulations. We will also use the reduced Planck constant, given by

$$h = \frac{h_0}{2\pi}$$

which is numerically given by

$$h \simeq 1.054\ 571\ 817 \times 10^{-34}$$

with everything being as usual in standard units.

To be more precise here, the point is that in quantum mechanics the constant which appears all the time is the reduced Planck constant, usually denoted $\hbar = h/2\pi$, with the original Planck constant being denoted h . However, the point is that, at least in what concerns us, we will heavily use for our quantum mechanics computations both this reduced Planck constant \hbar and the unnormalized trace of matrices tr . And the problem

is that, in quick handwriting, tr , \hbar often get confused, and when you compute you can't stop cursing, and you end up adopting the convention $h = h_0/2\pi$, as above.

Finally, let us mention that a very interesting continuation of Planck's work concerns the black body radiation of the early universe, with the microwave part of it, via a Doppler shift, still permeating the space that we live in. And with this phenomenon, called "cosmic microwave background", being at the origin of all modern cosmology.

4e. Exercises

Exercises.

Part II

Quantum electrodynamics

*In the midnight hour
She cried more, more, more
With a rebel yell
She cried more, more, more*

CHAPTER 5

Quantum mechanics

5a. Schrödinger equation

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

QUESTION 5.1. *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 initial 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 5.2. *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.*

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 5.1, in the following way:

QUESTION 5.3. *In the context of the hydrogen atom, assuming that the proton is fixed, what is the amplitude function $\psi_t(x)$ of the position of the electron e , at time t ,*

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

as function of an initial 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 5.4 (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 5.5. *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.

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

$$V = -\frac{e^2}{4\pi\epsilon_0} \cdot \frac{1}{r}$$

This potential is time-independent, and we have the following result:

THEOREM 5.6. *In the case of time-independent potentials V , including the Coulomb potential of the proton, the solutions of the Schrödinger equation*

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

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,

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

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

PROOF. By dividing by ψ , the equation becomes:

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

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

Moving ahead with theory, we can further build on Theorem 5.6, with a number of key observations on the time-independent Schrödinger equation, as follows:

THEOREM 5.7. *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 $\langle T \rangle$ are time-independent.*
- (2) *The time-independent Schrödinger equation can be written as $\hat{H}\phi = E\phi$, with $H = T + V$ being the total energy, of Hamiltonian.*
- (3) *For solutions of type $\psi = w_t\phi(x)$ we have $\langle H^k \rangle = E^k$ for any k . In particular we have $\langle H \rangle = E$, and the variance is $\langle H^2 \rangle - \langle H \rangle^2 = 0$.*

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

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

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

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

THEOREM 5.8. *The solutions of the Schrödinger equation with time-independent potential V appear as linear combinations of separated solutions*

$$\psi = \sum_n c_n e^{-iE_n t/\hbar} \phi_n$$

with the absolute values of the coefficients being given by

$$\langle H \rangle = \sum_n |c_n|^2 E_n$$

$|c_n|$ 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 [45]. \square

5b. 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 5.9. *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 $\{x_i\} \rightarrow \{y_i\}$, 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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 5.10. *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 of Δ in Theorem 5.9, 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. \square

5c. Spherical harmonics

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 5.11. *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 5.12. *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 than everything what we did so far, and so please be patient. We first have:

PROPOSITION 5.13. *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:

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

By differentiating at left, this equation becomes:

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

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

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

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

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

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

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

By plugging this data, our radial equation becomes:

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

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

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

But this is the Legendre equation, as stated. □

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

THEOREM 5.14. *The solutions of the Legendre equation, namely*

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

can be explicitly 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. □

In order to construct the solutions, let us start with something basic. We know 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, 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 5.15. *The orthonormal basis of $L^2[-1, 1]$ obtained by starting with the Weierstrass basis $\{x^l\}$, and doing Gram-Schmidt, is the family of polynomials $\{P_l\}$, with each P_l being of degree l , and with positive leading coefficient, subject to:*

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

These polynomials, called Legendre polynomials, satisfy the equation

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

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

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

which is called the Rodrigues formula for Legendre polynomials.

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

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

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

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

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

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

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

PROPOSITION 5.16. *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. \square

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 5.17. *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. \square

5d. Bohr energy

In order now to finish our study, 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 5.18. *The radial equation, written with $K = l(l + 1)$,*

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

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

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

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:

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

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

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

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. \square

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 chapters 1-3 above the story of the Coulomb constant K , which was eventually 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 5.19. *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 5.20. *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 5.21 (Schrödinger). *In the case of the hydrogen atom, where V is the Coulomb potential of the proton, the modified radial equation, which reads*

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

leads to the Bohr formula for allowed energies,

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

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, the idea being as follows:

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

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

In terms of $\alpha = \sqrt{-2mE}/\hbar$, 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 $p \rightarrow \infty$ 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 $p \rightarrow 0$ 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:

$$u' = p^l e^{-p} [(l+1-p)v + pv']$$

Differentiating a second time gives the following formula:

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

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 $j \gg 0$ 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:

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

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

$$\begin{aligned} K &= 8.988 \times 10^9 \quad , \quad e = 1.602 \times 10^{-19} \\ h &= 1.055 \times 10^{-34} \quad , \quad m = 9.109 \times 10^{-31} \end{aligned}$$

We obtain successively that we have the following formulae:

$$\begin{aligned}\frac{Ke^2}{h} &= \frac{8.988 \times 1.602^2}{1.055} \times \frac{10^9 \times 10^{-38}}{10^{-34}} = 2.186 \times 10^6 \\ \left(\frac{Ke^2}{h}\right)^2 &= 2.186^2 \times 10^{12} = 4.779 \times 10^{12} \\ \frac{m}{2} \left(\frac{Ke^2}{h}\right)^2 &= \frac{9.109 \times 4.779}{2} \times 10^{12-31} = 2.177 \times 10^{-18}\end{aligned}$$

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. \square

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

THEOREM 5.22. *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 5.21,

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

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:

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

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 5.21:

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

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

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

THEOREM 5.23. *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 $\rho_{nl}(r) = p^{l+1}e^{-pv}(p)/r$ 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 5.24. *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 5.21, 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 5.23 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. \square

In order to advance, we can use more mathematics, as follows:

PROPOSITION 5.25. *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:

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

The orthogonal basis obtained by applying Gram-Schmidt to the basis $\{x^q\}$ is then the basis formed by the Laguerre polynomials $\{L_q\}$, and this gives the results. \square

With the above result in hand, we can now improve Theorem 5.23, as follows:

THEOREM 5.26. *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 5.23 and Proposition 5.25, 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.

5e. Exercises

Exercises.

CHAPTER 6

Fine structure

6a. Fine structure

What is next? All sorts of corrections to the solution that we found, due to various phenomena that we neglected in our computations, or rather in our modelling of the problem, which can be both of electric and relativistic nature. But before that, let us first enjoy what we found. As a first conclusion, of 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, in our study we dismissed several times all sorts of solutions, on various physical grounds. But, at a more advanced level, we have:

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. But we will not get into this, here.

Back to work now, 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 5 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 an estimate for Ke^2/h , from chapter 5. □

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 $K = 1/(4\pi\epsilon_0)$ and $h = h_0/2\pi$, we can write this constant as:

$$\alpha = \frac{e^2}{2\epsilon_0 h_0 c}$$

Finally, let us record the complete official data for α and its inverse α^{-1} :

$$\alpha = 0.007\ 297\ 352\ 5693(11)$$

$$\alpha^{-1} = 137.035\ 999\ 084(21)$$

As a final comment here, all this lengthy discussion about α might sound a bit like mania, or mysticism. But wait for it. Sometimes soon α will be part of your life.

Moving ahead now with corrections to Theorem 6.3, we will be very brief, and for further details, we refer as usual to our favorite books, Feynman [39], Griffiths [45] and Weinberg [90]. We first have the following result, which is something non-trivial:

THEOREM 6.4. *There is a relativistic correction to be made to the Bohr energy E_n of the state ϕ_{nlm} , depending on the quantum number l , given by*

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

coming by replacing the kinetic energy by the relativistic kinetic energy.

PROOF. According to Einstein, the relativistic kinetic energy is given by:

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

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

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

THEOREM 6.5. *There is a spin-related correction to be made to the Bohr energy E_n of the state ϕ_{nlm} , depending on the number $j = l \pm 1/2$, given by*

$$\mathcal{E}_{nj} = -\frac{\alpha^2 E_n}{n^2} \cdot \frac{n(j-l)}{(l+1/2)(j+1/2)}$$

coming from the torque of the proton on the magnetic moment of the electron.

PROOF. As we will explain later, 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 [39], Griffiths [45], Weinberg [90] for details. Now with this in hand, we obtain:

THEOREM 6.6. *The energy levels of the hydrogen atom, taking into account the fine structure coming from the relativistic and spin-related correction, are given by*

$$E_{nj} = E_n \left[1 + \frac{\alpha^2}{n^2} \left(\frac{n}{j + 1/2} - \frac{3}{4} \right) \right]$$

with $j = l \pm 1/2$ being as above, and with α being the fine structure constant.

PROOF. We have the following computation, based on the above formulae:

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

Thus the corrected formula of the energy is as follows:

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

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

Summarizing, quantum mechanics is more complicated than what originally appears from Schrödinger's solution of the hydrogen atom. Which was something quite complicated too, we must admit that. And the story is not over here, because on top of the above fine structure correction, which is of order α^2 , we have afterwards the Lamb shift, which is an order α^3 correction, then the hyperfine splitting, and more.

As usual, we refer to Feynman [39], Griffiths [45], Weinberg [90] 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 [40], who managed to find a global way of viewing all the phenomena that can appear, corresponding to an infinite series in α .

6b.

6c.

6d.

6e. Exercises

CHAPTER 7

Feynman diagrams

7a.

7b.

7c.

7d.

7e. Exercises

CHAPTER 8

Renormalization

8a.

8b.

8c.

8d.

8e. Exercises

Part III

Twisted electrodynamics

*The time has come
To say fair's fair
To pay the rent
To pay our share*

CHAPTER 9

9a.

9b.

9c.

9d.

9e. Exercises

CHAPTER 10

10a.

10b.

10c.

10d.

10e. Exercises

CHAPTER 11

11a.

11b.

11c.

11d.

11e. Exercises

CHAPTER 12

12a.

12b.

12c.

12d.

12e. Exercises

Part IV

Free electrodynamics

*It has to start somewhere
It has to start sometime
What better place than here
What better time than now*

CHAPTER 13

13a.

13b.

13c.

13d.

13e. Exercises

CHAPTER 14

14a.

14b.

14c.

14d.

14e. Exercises

CHAPTER 15

15a.

15b.

15c.

15d.

15e. Exercises

CHAPTER 16

16a.

16b.

16c.

16d.

16e. Exercises

Bibliography

- [1] V.I. Arnold, Ordinary differential equations, Springer (1973).
- [2] V.I. Arnold, Lectures on partial differential equations, Springer (1997).
- [3] V.I. Arnold and B.A. Khesin, Topological methods in hydrodynamics, Springer (1998).
- [4] M.F. Atiyah, The geometry and physics of knots, Cambridge Univ. Press (1990).
- [5] T. Banica, Introduction to modern physics (2023).
- [6] T. Banica, Free differential equations (2023).
- [7] T. Banica, Invitation to lattice models (2023).
- [8] R.J. Baxter, Exactly solved models in statistical mechanics, Academic Press (1982).
- [9] I. Bengtsson and K. Życzkowski, Geometry of quantum states, Cambridge Univ. Press (2006).
- [10] H. Bercovici and V. Pata, Stable laws and domains of attraction in free probability theory, *Ann. of Math.* **149** (1999), 1023–1060.
- [11] J. Bhowmick, F. D’Andrea and L. Dabrowski, Quantum isometries of the finite noncommutative geometry of the standard model, *Comm. Math. Phys.* **307** (2011), 101–131.
- [12] J. Bhowmick, F. D’Andrea, B. Das and L. Dabrowski, Quantum gauge symmetries in noncommutative geometry, *J. Noncommut. Geom.* **8** (2014), 433–471.
- [13] R. Brauer, On algebras which are connected with the semisimple continuous groups, *Ann. of Math.* **38** (1937), 857–872.
- [14] S.M. Carroll, Spacetime and geometry, Cambridge Univ. Press (2004).
- [15] A.H. Chamseddine and A. Connes, The spectral action principle, *Comm. Math. Phys.* **186** (1997), 731–750.
- [16] A.H. Chamseddine and A. Connes, Why the standard model, *J. Geom. Phys.* **58** (2008), 38–47.
- [17] V. Chari and A. Pressley, A guide to quantum groups, Cambridge Univ. Press (1994).
- [18] S. Coleman, Quantum field theory, World Scientific (2019).
- [19] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic groups, *Comm. Math. Phys.* **264** (2006), 773–795.
- [20] A. Connes, Noncommutative geometry, Academic Press (1994).
- [21] A. Connes, On the spectral characterization of manifolds, *J. Noncommut. Geom.* **7** (2013), 1–82.
- [22] A. Connes and M. Marcolli, Noncommutative geometry, quantum fields and motives, AMS (2008).
- [23] W.N. Cottingham and D.A. Greenwood, An introduction to the standard model of particle physics, Cambridge Univ. Press (2012).

- [24] B. Das, U. Franz and X. Wang, Invariant Markov semigroups on quantum homogeneous spaces, *J. Noncommut. Geom.* **15** (2021), 531–580.
- [25] P.A. Davidson, Introduction to magnetohydrodynamics, Cambridge Univ. Press (2001).
- [26] P. Di Francesco, Meander determinants, *Comm. Math. Phys.* **191** (1998), 543–583.
- [27] P. Di Francesco, P. Mathieu and D. Sénéchal, Conformal field theory, Springer (1996).
- [28] P.A.M. Dirac, Principles of quantum mechanics, Oxford Univ. Press (1930).
- [29] M.P. do Carmo, Riemannian geometry, Birkhäuser (1992).
- [30] S. Dodelson, Modern cosmology, Academic Press (2003).
- [31] R. Durrett, Probability: theory and examples, Cambridge Univ. Press (1990).
- [32] A. Einstein, Relativity: the special and the general theory, Dover (1916).
- [33] L.C. Evans, Partial differential equations, AMS (1998).
- [34] L.D. Faddeev and L. A. Takhtajan, Hamiltonian methods in the theory of solitons, Springer (2007).
- [35] W. Feller, An introduction to probability theory and its applications, Wiley (1950).
- [36] E. Fermi, Thermodynamics, Dover (1937).
- [37] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics I: mainly mechanics, radiation and heat, Caltech (1963).
- [38] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics II: mainly electromagnetism and matter, Caltech (1964).
- [39] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics III: quantum mechanics, Caltech (1966).
- [40] R.P. Feynman and A.R. Hibbs, Quantum mechanics and path integrals, Dover (1965).
- [41] D. Goswami and J. Bhowmick, Quantum isometry groups, Springer (2016).
- [42] J.M. Gracia-Bondía, J.C. Várilly and H. Figueroa, Elements of noncommutative geometry, Birkhäuser (2001).
- [43] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory, Cambridge Univ. Press (2012).
- [44] D.J. Griffiths, Introduction to electrodynamics, Cambridge Univ. Press (2017).
- [45] D.J. Griffiths and D.F. Schroeter, Introduction to quantum mechanics, Cambridge Univ. Press (2018).
- [46] D.J. Griffiths, Introduction to elementary particles, Wiley (2020).
- [47] D.J. Griffiths, Revolutions in twentieth-century physics, Cambridge Univ. Press (2012).
- [48] K. Huang, Introduction to statistical physics, CRC Press (2001).
- [49] K. Huang, Quantum field theory, Wiley (1998).
- [50] K. Huang, Quarks, leptons and gauge fields, World Scientific (1982).
- [51] K. Huang, Fundamental forces of nature, World Scientific (2007).
- [52] C. Itzykson and J.B. Zuber, Quantum field theory, Dover (1980).
- [53] J.D. Jackson, Classical electrodynamics, Wiley (1962).
- [54] V.F.R. Jones, Index for subfactors, *Invent. Math.* **72** (1983), 1–25.

- [55] V.F.R. Jones, On knot invariants related to some statistical mechanical models, *Pacific J. Math.* **137** (1989), 311–334.
- [56] V.F.R. Jones, Planar algebras I, preprint (1999).
- [57] L.P. Kadanoff, Statistical physics: statics, dynamics and renormalization, World Scientific (2000).
- [58] T. Kibble and F.H. Berkshire, Classical mechanics, Imperial College Press (1966).
- [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] G. Landi, An introduction to noncommutative spaces and their geometry, Springer (1997).
- [66] Y.I. Manin, Quantum groups and noncommutative geometry, Springer (2018).
- [67] V.A. Marchenko and L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, *Mat. Sb.* **72** (1967), 507–536.
- [68] M. Marcolli, Noncommutative cosmology, World Scientific (2018).
- [69] M.L. Mehta, Random matrices, Elsevier (2004).
- [70] J. Nash, The imbedding problem for Riemannian manifolds, *Ann. of Math.* **63** (1956), 20–63.
- [71] M.A. Nielsen and I.L. Chuang, Quantum computation and quantum information, Cambridge Univ. Press (2000).
- [72] W.K.H. Panofsky and M. Phillips, Classical electricity and magnetism, Dover (1955).
- [73] A. Peres, Quantum theory: concepts and methods, Kluwer (1993).
- [74] M. Peskin and D.V. Schroeder, An introduction to quantum field theory, CRC press (1995).
- [75] J. Polchinski, String theory, Cambridge Univ. Press (2001).
- [76] E.M. Purcell and D.J. Morin, Electricity and magnetism, Cambridge Univ. Press (1965).
- [77] J.J. Sakurai and J. Napolitano, Modern quantum mechanics, Cambridge Univ. Press (1985).
- [78] M. Schwartz, Principles of electrodynamics, Dover (1972).
- [79] J. Schwinger, Einstein’s legacy: the unity of space and time, Dover (1986).
- [80] J. Schwinger, L.L. DeRaad Jr., K.A. Milton and W.Y. Tsai, Classical electrodynamics, CRC Press (1998).
- [81] J. Schwinger and B.H. Englert, Quantum mechanics: symbolism of atomic measurements, Springer (2001).
- [82] J. Smit, Introduction to quantum fields on a lattice, Cambridge Univ. Press (2002).
- [83] M.A. Srednicki, Quantum field theory, Cambridge Univ. Press (1980).

- [84] J.R. Taylor, *Classical mechanics*, Univ. Science Books (2003).
- [85] J.S. Townsend, *A modern approach to quantum mechanics*, Univ. Science Books (1992).
- [86] D.V. Voiculescu, K.J. Dykema and A. Nica, *Free random variables*, AMS (1992).
- [87] J. von Neumann, *Mathematical foundations of quantum mechanics*, Princeton Univ. Press (1955).
- [88] S. Wang, Quantum symmetry groups of finite spaces, *Comm. Math. Phys.* **195** (1998), 195–211.
- [89] S. Weinberg, *Foundations of modern physics*, Cambridge Univ. Press (2011).
- [90] S. Weinberg, *Lectures on quantum mechanics*, Cambridge Univ. Press (2012).
- [91] S. Weinberg, *The theory of quantum fields*, Cambridge Univ. Press (1995).
- [92] S. Weinberg, *Cosmology*, Oxford Univ. Press (2008).
- [93] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank, *J. Math. Phys.* **19** (1978), 999–1001.
- [94] H. Weyl, *The theory of groups and quantum mechanics*, Princeton Univ. Press (1931).
- [95] H. Weyl, *The classical groups: their invariants and representations*, Princeton Univ. Press (1939).
- [96] H. Weyl, *Space, time, matter*, Princeton Univ. Press (1918).
- [97] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, *Ann. of Math.* **62** (1955), 548–564.
- [98] E. Witten, Quantum field theory and the Jones polynomial, *Comm. Math. Phys.* **121** (1989), 351–399.
- [99] S.L. Woronowicz, Compact matrix pseudogroups, *Comm. Math. Phys.* **111** (1987), 613–665.
- [100] B. Zwiebach, *A first course in string theory*, Cambridge Univ. Press (2004).