The joys of relativity theory

Teo Banica

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

2010 *Mathematics Subject Classification.* 83A05

Key words and phrases. Special relativity, General relativity

Abstract. This is an introduction to Einstein's relativity theory, both special and general, assuming basic calculus known. We carefully explain the foundations of classical physics, and the various experiments leading to special relativity, and then we present the basic results of the theory. In relation with general relativity, we pay particular attention to the explanation of the various mathematical notions involved, such as the differential manifolds, and calculus on them, and then we present the basic results of the theory. Finally, we discuss some applications to compact stars and black holes.

Preface

Classical mechanics, and substantial parts of electromagnetism, thermodynamics, and even quantum mechanics, and applications of these, to basic questions in chemistry, engineering and so on, are largely "linear". The variables there, be them refering to space, time or other quantities, are allowed to be arbitrarily big or arbitrarily small, and are also allowed to move, with the same ease, in the positive direction and in the negative direction. And with all this being certainly pleasant, and in tune with our intuition.

At a more advanced level, however, things turn complicated, for several reasons. Historically, the first discovery was that temperature is bounded from below, $T > 0$. Then came the discovery that speed is bounded from above, $v < c$. This latter discovery was of remarkable impact, with $v = d/t$ forcing Einstein and others to rethink previously almighty notions like distance d, time t, and also mass m, momentum p, energy E and so on, with the conclusion that nothing escapes from the fallout of $v < c$, and everything is bounded and curved, instead of being linear as previously thought.

On top of this, roughly at the same time, end of the 19th century and beginning of the 20th century, appeared the idea of quantization. Again things here came from thermodynamics, with Planck's finding that heat is bounded from below, $\Delta T > \varepsilon$. As with $v < c$, this was something of remarkable impact, eventually leading, via quantum mechanics, to the conclusion that poor v, d, t, m, p, E, T, already curved by relativity, must in addition evolve in discrete increments, called quanta. So, end of the world as we know it, things are both curved and discrete, and we will have to live with that.

Fortunately, these deep findings do not affect much the bulk of useful science and engineering, which are based anyway on good old methods from the 18th and 19th centuries, or older. The reasons for this come from the fact that both the relativistic and quantum corrections are truly tiny, in regards with the questions that we usually deal with, in the real life surrounding us, where all the variables are not too small, but not too big either. However, at smaller or bigger scales both relativity and quantum theory must be taken into account, and at the large end, that of stars, galaxies, and the universe itself, this has led to a lot of interesting science, developed all over the 20th century.

The present book is an introduction to $v < c$ and relativity, and so to half of the problem, so to say. We have tried to keep it as nice and pleasant as possible. Again, the main problem with relativity, and in fact with quantum mechanics too, is not that of getting familiar with the mathematics and formulae, which is something technical that can certainly be done, if you have to, for your job or other, but rather to "agree" with the theory, meaning to really love that theory, from the bottom of your heart. So, in the hope that you will find this book useful, and get to love these damn things. At least it worked for me, after writing this book I find relativity really cool.

More in detail now, we will carefully explain the foundations of classical physics, and the various experiments leading to special relativity, and then we will present the basic results of the theory. In relation with general relativity, we will pay particular attention to the explanation of the various mathematical notions involved, such as the differential manifolds, and calculus on them, and then we will present the basic results of the theory. Finally, we will discuss some applications to compact stars and black holes.

I would like to thank Jean-Marc Schlenker and my geometer colleagues, on numerous occasions they tried to convert me to their faith in modern science and relativity, and with a bit of delay, I think they have succeeded. Many thanks go as well to my cats. Relativity is a bit hard to trust in, but having some relativistic little friends around, with undefined positions and speeds, and often travelling via wormholes, helps a lot.

Contents

Part I

Special relativity

Purity attack
Purity rock
Purity on top
And purity non-stop

CHAPTER 1

Inertial frames

1a. Speed of light

You probably heard about Einstein, and his relativity theory, leading to $E = mc^2$ and many other things. Einstein wrote a book [28] about all this. That book is quite compact, around 100-150 pages depending on editions, public domain and being printed and sold for cheap, typically around 10 euros or dollars, and is very enjoyable.

Get a copy of that book, and read it. Not only you will learn many things about relativity, from the one who has invented it, but you will learn more about Einstein himself too. You will be surprised, the book is very readable, written with a lot of humanity and common sense, to the point that you soon forget that you're reading physics, and even start wondering if Einstein was really a genius, or just some random guy passioned by physics, and writing a book about things that he found, in his spare time.

So, that would be my first advice, learn a bit about Einstein and his work, directly from himself. And why not about other known mathematicians, physicists and other scientists, directly from themselves too. Mathematics, physics and science are simple and beautiful, always believe in this, and don't let anyone make you think otherwise.

Getting to the present book now, our aim here is to explain Einstein's theory in [28], from a modern perspective. We will insist on one hand on lots of formulae and mathematics, usually in relation with basic calculus, but sometimes in relation with tricky differential geometry too, and with this taking a lot of space, and making the present book considerably long. Also, as a main application of relativity, we will talk at the end about cosmology, compact stars and black holes, following a lot of science developed in the 20th century, after Einstein's book [28]. All in all, very standard way of approaching relativity, as of now, early 21th century, and in the hope that you will enjoy all this.

In order to get started, we need a so-called wormhole, which is a device which can be purchased for cheap on the internet, but of course watch out for fakes. So, travelling together back in time, around 1900, let us ask ourselves the following question:

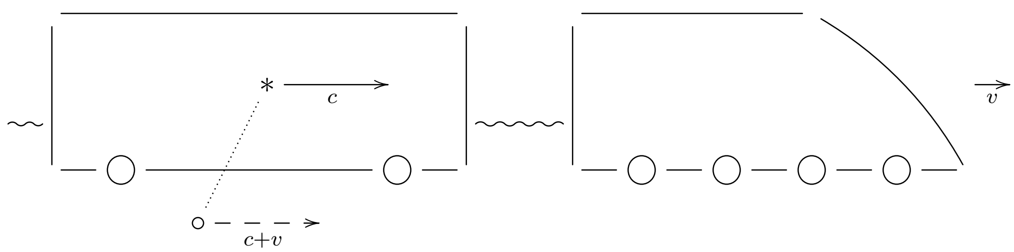
Question 1.1. *Is classical mechanics correct?*

Normally yes, why should it be wrong. However, based on various experiments, that we will describe in a moment, and a bit of abstract thinking too, Einstein came upon the following principles, which obviously contradict classical mechanics:

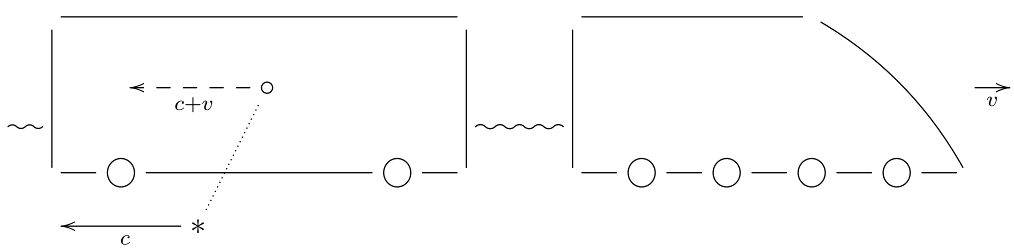
FACT 1.2 (Einstein principles). *The following happen:*

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

All this needs of course some details and explanations, and we will come to this, in a moment. However, before that, let us see what the contradiction with classical mechanics is. This basically comes from (2), which is the tricky point in the above, because if we assume that we have a train, running in vacuum at speed $v > 0$, and someone on board lights a flashlight $*$ towards the locomotive, then an observer $\circ$ on the ground will see the light travelling at speed $c + v > c$, which is a contradiction:



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



With this understood, let us get now into various technical details, regarding Fact 1.2. We must talk a bit about light, about vacuum, and about inertial frames. Regarding light, let us agree that this is the light that we all know, and more of course can be said, but no hurry here, and we will be back to this later, in chapter 8 below.

Regarding vacuum and non-vacuum, these are the vacuum and non-vacuum that we all know, and we will not comment about this either. Let us mention however that light travels indeed at speed $v < c$ in non-vacuum, with the speed lost coming from the heating of the material, and with the vacuum being of course impossible to heat, for the simple reason that there is nothing there that we can heat. As before, this sort of intuitive knowledge will do, and more on this later, with details, in chapter 8 below.

More concretely now, in relation with Fact 1.2 (1), according to various experiments and regulations, we have a formula for the light speed in vacuum c, as follows:

FACT 1.3. *The speed of light in vaccum is*

$$c = 299\,792\,458 \text{ m/s}$$

with this being an exact formula.

This is a figure which comes out of experiments, and with $c \simeq 3 \times 10^9$ m/s being actually something quite old. As an interesting feature, as mentioned, the above formula is exact, due to the well-known mess with regulating meters, seconds and other units. To be more precise, there is a rock-solid definition for the second, based on atomic clocks, but not for the meter, and by a recent decision the above formula was declared exact, and any further changes in the observed value of c will be blamed on the meter.

Be said in passing, would such a wise decision have been taken much earlier, the speed of light would be $c = 3 \times 10^9$ m/s, exactly. But now it is too late. Finally, all these issues do not bother much theoretical physicists, who use tailored units depending on the precise problems they are working on, with around 50 orders of magnitude on the menu, and who usually, when dealing often with c, arrange things as to take $c = 1$.

Actually, speaking $c = 1$, we will often use this convention, as follows:

CONVENTION 1.4. *In the context of computations with high speeds, comparable to c, which do not involve other physics and units than those for length and time, we will use the convention $c = 1$, coming from a linear rescaling of length, or of time, or both.*

Observe the uncertainty at the end, regarding the exact nature of the rescaling that we are doing. This is not exactly something very orthodox in the context of regular mathematics and physics, but is quite commonplace in the context of pure relativity theory, and we will often use this convention, which is quite practical for the production and understanding of various math formulae. In case this uncertainty bothers you, you can imagine, for being fully rigorous, that one of the following happens:

(1) You are the Sun, and $299,792,458$ human meters appear to you as 1 meter.

(2) You are a tiger, and 1 human second appears to you as $1/299,792,458$ seconds.

(3) You are a green alien, and in your alien world, only speeds, close to 1, exist.

But more on this later, we are not there, yet. Back now to Fact 1.2, there are still several things to be clarified there, and more specifically we have to talk about:

(1) Inertial frames. This is something quite technical, and we will come to this in a moment, with full details. For the moment, let us simply agree that Fact 1.2 (2) says that "c is about the same for all reasonable observers", with reasonable meaning a bit of everything, just excluding situations like zombies riding on elementary particles.

(2) Experiments, evidence. Again, this is something a bit technical, with the evidence, coming from a key experiment of Fizeau, then a subsequent experiment by Michelson-Morley, and then various findings involving Maxwell and Lorentz, supporting what Einstein built starting from Fact 1.2, rather than Fact 1.2 itself. More on this later.

Long story short, in order to close this preliminary opening section of this book, and explain what relativity theory is about, let us agree that Fact 1.2 logically makes sense, and is physically correct too. But, with this agreed upon, as explained above, this fact tells us that classical mechanics is wrong as stated, and needs a correction.

But, how to do this correction? Mathematically speaking, we must deal here with something of type $c + v = c$, and if there is a number made for that, that is $c = \infty$. But Fact 1.2 (1) clearly says that $c < \infty$, so no good. We must come up with something more complicated, and thinking and thinking about all this, which is what Einstein did, following his discovery, leads to the following conclusion:

Conclusion 1.5 (Einstein). *Things in classical mechanics are a bit curved at the $v = \infty$ end, leading to $v < c$. In order to fix classical mechanics, we must add a bit of curvature in all formulae, allowing at the same time for $c + v = c$, and $c < \infty$.*

We will go on this path, following Einstein, in a moment, and we will discover that the curvature of the world that we're living in coming from speed, at the $v = \infty$ end, forces in fact many other familiar things, such as the distances d, masses m, energies E, and even the time t itself, to be curved as well. And with this being no longer something speculative, but just truths, coming from basic mathematics, based on Fact 1.2.

Excited about this? In short, we will be embarking on a total destruction process, dealing with the ∞ end of everything that we know, and always took for granted. Let us mention that other familiar things like the temperature T will not be spared either, and we will talk about this later, when doing a bit of thermodynamics.

Importantly, all these things that we've been talking about, namely speed v, distance d, mass m, time t, energy E and temperature T, and perhaps some other too, will turn to be curved at their 0 end as well, due to all sorts of other phenomena, which are of more complicated, quantum mechanical nature. So, anticipating a bit, and including a number of things that we don't know yet, let us upgrade Conclusion 1.5 into:

FACT 1.6. *Everything in life is a bit curved at the ∞ end, and at the 0 end too.*

With this being more than enough as a philosophical conclusion, normally guaranteeing a good place in an asylum, let us get now back to work.

1b. Kepler and Newton

Before going ahead with relativity, in order to fully understand what Fact 1.2 says, we still have to discuss the inertial frames. And, there is some work to be done here. Indeed, this notion is something quite subtle, and not necessarily very intuitive, the problem being for instance that our beloved Earth is not an inertial frame, and nor is the Moon, or Jupiter, or the Sun, or the Halley comet, or Sirus, Betelgeuse or Rigel.

It is possible to talk about inertial frames abstractly, just in the context of the laws of motion, but we will not do this, because this will make us miss the whole point. So, let us go instead to the usual, gravitational context of classical mechanics.

You probably know a bit about classical mechanics, but this is beautiful physics, always good to remember. As a starting point, following Kepler, we have:

FACT 1.7 (Kepler laws). *The following happen:*

(1) *The planetary orbits are elliptical, with the Sun at a focus.*
(2) *The radius vector from the Sun to a planet sweeps equal areas in equal times.*
(3) *The ratio of the square of the period of revolution and the cube of the ellipse semimajor axis is the same for all planets.*

The Kepler laws were a big achievement when they came out, found by Kepler by browsing through a massive amount of astronomical data, accumulated over centuries, and even millenia. However, it is possible to do even better. Following Newton, based on Kepler, plus once again on massive amounts of astronomical data, we have:

FACT 1.8 (Newton principles). *The force of attraction between two bodies of masses m_1, m_2, having distance $d > 0$ between them, is given by the formula*

$$||F|| = G \cdot \frac{m_1 m_2}{d^2}$$

where $G \simeq 6.674 \times 10^{-11}$. This force alters the trajectory of one body with respect to another, and more specifically the position x, speed v, and acceleration a, via

$$F = ma \quad , \quad a = \dot{v} \quad , \quad v = \dot{x}$$

where the dot denotes the derivative with respect to time, with this meaning the infinitesimal rate of change of the given quantity, with respect to time.

All this might look a bit complicated, but the point is that we have the following result, due to Newton, with contributions of course by Kepler, and by the ancient Greeks too, which is the pride of mathematics, physics, and human knowledge in general:

Theorem 1.9. *The following happen:*

(1) *Planets and other celestial bodies move around the Sun on conics, that is, curves given by $P(x,y)=0$, with $P\in\mathbb{R}[x,y]$ being of degree 2.*
(2) *The conics are the curves which appear by cutting a 2-sided cone with a plane, and can be classified into ellipses, parabolas and hyperbolas.*

In addition, the movements are subject as well to the Kepler 2 and 3 laws.

Proof. The idea here is that (1) is something tough, due to Newton, and (2), due to the ancient Greeks, is elementary. Let us mention too that the above statement is a bit informal, with the 3 viewpoints on the conics, coming from gravity, cutting cones and classification, agreeing in the non-degenerate case, $\deg P=2$, modulo some normalizations, and with some disagreements in the degenerate case, $\deg P\leq 1$. But more on this later, towards the end of the proof, we will discuss all this at that time.

(1) The starting point is the main formula in Fact 1.8, telling us that the force of attraction between two bodies of masses M,m, at distance $d>0$, is given by:

$$||F||=G\cdot\frac{Mm}{d^2}$$

Now by assuming that M is fixed at $0\in\mathbb{R}^3$, the force exterted on m positioned at $x\in\mathbb{R}^3$, regarded as a vector $F\in\mathbb{R}^3$, is given by the following formula:

$$F=-||F||\cdot\frac{x}{||x||}=-\frac{GMm}{||x||^2}\cdot\frac{x}{||x||}=-\frac{GMmx}{||x||^3}$$

But $F=ma=m\ddot{x}$, with $a=\ddot{x}$ being the acceleration, second derivative of the position, so the equation of motion of m, assuming that M is fixed at 0, is:

$$\ddot{x}=-\frac{GMx}{||x||^3}$$

Obviously, the problem happens in 2 dimensions, and you can even find, as an exercise, a formal proof of that, based on the above equation, if your really want to. Now here the most convenient is to use standard x,y coordinates, and denote our point as $z=(x,y)$. With this change made, and by setting $K=GM$, the equation of motion becomes:

$$\ddot{z}=-\frac{Kz}{||z||^3}$$

(2) The idea now is that the problem can be solved via some calculus. Let us write indeed our vector $z=(x,y)$ in polar coordinates, as follows:

$$x=r\cos\theta$$

$$y=r\sin\theta$$

We have then $||z|| = r$, and our equation of motion becomes:

$$\ddot{z} = -\frac{Kz}{r^3}$$

Let us differentiate now x, y. By using the standard calculus rules, we have:

$$\dot{x} = \dot{r}\cos\theta - r\sin\theta \cdot \dot{\theta}$$

$$\dot{y} = \dot{r}\sin\theta + r\cos\theta \cdot \dot{\theta}$$

Differentiating one more time gives the following formulae:

$$\ddot{x} = \ddot{r}\cos\theta - 2\dot{r}\sin\theta \cdot \dot{\theta} - r\cos\theta \cdot \dot{\theta}^2 - r\sin\theta \cdot \ddot{\theta}$$

$$\ddot{y} = \ddot{r}\sin\theta + 2\dot{r}\cos\theta \cdot \dot{\theta} - r\sin\theta \cdot \dot{\theta}^2 + r\cos\theta \cdot \ddot{\theta}$$

Consider now the following two quantities, appearing as coefficients in the above:

$$a = \ddot{r} - r\dot{\theta}^2 \quad , \quad b = 2\dot{r}\dot{\theta} + r\ddot{\theta}$$

In terms of these quantities, our second derivative formulae read:

$$\ddot{x} = a\cos\theta - b\sin\theta$$

$$\ddot{y} = a\sin\theta + b\cos\theta$$

(3) We can now solve the equation of motion from (1). Indeed, with the formulae that we found for $\ddot{x}, \ddot{y}$, our equation of motion takes the following form:

$$a\cos\theta - b\sin\theta = -\frac{K}{r^2}\cos\theta$$

$$a\sin\theta + b\cos\theta = -\frac{K}{r^2}\sin\theta$$

But these two formulae can be written in the following way:

$$\left(a + \frac{K}{r^2}\right)\cos\theta = b\sin\theta$$

$$\left(a + \frac{K}{r^2}\right)\sin\theta = -b\cos\theta$$

By making now the product, and assuming that we are in a non-degenerate case, where the angle θ varies indeed, we obtain by positivity that we must have:

$$a + \frac{K}{r^2} = b = 0$$

(4) Let us first examine the second equation, $b = 0$. This can be solved as follows:

$$\begin{aligned} b=0 &\iff 2\dot{r}\dot{\theta}+r\ddot{\theta}=0\\ &\iff \frac{\ddot{\theta}}{\dot{\theta}}=-2\frac{\dot{r}}{r}\\ &\iff (\log\dot{\theta})'=(-2\log r)'\\ &\iff \log\dot{\theta}=-2\log r+c\\ &\iff \dot{\theta}=\frac{\lambda}{r^2}\end{aligned}$$

As for the first equation the we found, namely $a + K/r^2 = 0$, this becomes:

$$\ddot{r}-\frac{\lambda^2}{r^3}+\frac{K}{r^2}=0$$

As a conclusion to all this, in polar coordinates, $x = r\cos\theta$, $y = r\sin\theta$, our equations of motion are as follows, with λ being a constant, not depending on t:

$$\ddot{r}=\frac{\lambda^2}{r^3}-\frac{K}{r^2}\quad,\quad \dot{\theta}=\frac{\lambda}{r^2}$$

Even better now, let us introduce a new constant K, as follows:

$$K=\frac{\lambda^2}{c}$$

With this convention, our equations above simply read:

$$\ddot{r}=\frac{\lambda^2}{r^2}\left(\frac{1}{r}-\frac{1}{c}\right)\quad,\quad \dot{\theta}=\frac{\lambda}{r^2}$$

(5) In order to study the first equation, we use a trick. Let us write:

$$r(t)=\frac{1}{f(\theta(t))}$$

Abbreviated, and by reminding that f takes $\theta = \theta(t)$ as variable, this reads:

$$r=\frac{1}{f}$$

With the convention that dots mean as usual derivatives with respect to t, and that the primes will denote derivatives with respect to $\theta = \theta(t)$, we have:

$$\dot{r}=-\frac{f'\dot{\theta}}{f^2}=-\frac{f'}{f^2}\cdot\frac{\lambda}{r^2}=-\lambda f'$$

By differentiating one more time with respect to t, we obtain:

$$\ddot{r}=-\lambda f''\dot{\theta}=-\lambda f''\cdot\frac{\lambda}{r^2}=-\frac{\lambda^2}{r^2}f''$$

On the other hand, our equation for $\ddot{r}$ found in (4) above reads:

$$\ddot{r}=\frac{\lambda^2}{r^2}\left(\frac{1}{r}-\frac{1}{c}\right)=\frac{\lambda^2}{r^2}\left(f-\frac{1}{c}\right)$$

Thus, in terms of $f=1/r$ as above, our equation for $\ddot{r}$ simply reads:

$$f''+f=\frac{1}{c}$$

But this latter equation is elementary to solve. Indeed, both functions $\cos t,\sin t$ satisfy $g"+g=0$, so any linear combination of them satisfies as well this equation. But the solutions of $f''+f=1/c$ being those of $g''+g=0$ shifted by $1/c$, we obtain:

$$f=\frac{1+\varepsilon\cos\theta+\delta\sin\theta}{c}$$

Now by inverting, we obtain the following formula:

$$r=\frac{c}{1+\varepsilon\cos\theta+\delta\sin\theta}$$

(6) But this leads to the conclusion that the trajectory is a conic. Indeed, in terms of the parameter θ, the formulae of the coordinates are:

$$x=\frac{c\cos\theta}{1+\varepsilon\cos\theta+\delta\sin\theta}$$

$$y=\frac{c\sin\theta}{1+\varepsilon\cos\theta+\delta\sin\theta}$$

Now observe that these two functions x,y satisfy the following formula:

$$x^2+y^2=\frac{c^2(\cos^2\theta+\sin^2\theta)}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2}=\frac{c^2}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2}$$

On the other hand, these two functions satisfy as well the following formula:

$$\begin{aligned}(\varepsilon x+\delta y-c)^2 &= \frac{c^2\big(\varepsilon\cos\theta+\delta\sin\theta-(1+\varepsilon\cos\theta+\delta\sin\theta)\big)^2}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2}\\ &= \frac{c^2}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2}\end{aligned}$$

We conclude that our coordinates x,y satisfy the following equation:

$$x^2+y^2=(\varepsilon x+\delta y-c)^2$$

But what we have here is an equation of a conic, and this ends the proof of the first assertion. The last assertion, regarding the Kepler 2 and 3 laws, can be proved as well, by further studying the above formulae, and we will be back to this later.

(7) The classification of the conics, going back to the ancient Greeks, is standard. Consider indeed one of these conics:

$$C = \left\{(x,y)\in\mathbb{R}^2\Big|P(x,y)=0,\deg P\leq 2\right\}$$

By doing some suitable manipulations on the degree 2 polynomial $P\in\mathbb{R}[x,y]$, up to affine transformations of the curve, we can have this curve written in some simple, "standard" form, with standard depending a bit on you, matter of taste. But this standard form can only lead to the 3 cases in the statement, namely ellipses, parabolas and hyperbolas, up to degeneration, with the degenerate cases being the lines, double lines, points, empty set, and $\mathbb{R}^2$ itself, basically appearing when $\deg P\leq 1$.

(8) The fact that the conics appear by cutting a 2-sided cone with a plane is also elementary, and also known since the ancient Greeks. A first proof is by doing some abstract algebra, and verifying that the cut must be indeed a curve of degree 2. A second proof is by computing the cut in the various cases that might appear, depending on the angle of the plane with respect to the cone, with this leading to the curves found in (7), namely ellipses, parabolas and hyperbolas, up to degeneration. □

Still with me, I hope, after all these computations. For further applications, here is a sort of "best of" the formulae found in the proof of Theorem 1.9:

THEOREM 1.10. *In the context of a 2-body problem, with M fixed at 0, and m starting its movement from Ox, the equation of motion of m, namely*

$$\ddot{z} = -\frac{Kz}{||z||^3}$$

with $K=GM$, and $z=(x,y)$, becomes in polar coordinates, $x=r\cos\theta$, $y=r\sin\theta$,

$$\ddot{r} = \frac{\lambda^2}{r^2}\left(\frac{1}{r}-\frac{1}{c}\right) \quad , \quad \dot{\theta} = \frac{\lambda}{r^2}$$

for some $\lambda,c\in\mathbb{R}$, related by $\lambda^2=Kc$. The value of r in terms of θ is given by

$$r = \frac{c}{1+\varepsilon\cos\theta+\delta\sin\theta}$$

for some $\varepsilon,\delta\in\mathbb{R}$. At the level of the affine coordinates x,y, this means

$$x = \frac{c\cos\theta}{1+\varepsilon\cos\theta+\delta\sin\theta} \quad , \quad y = \frac{c\sin\theta}{1+\varepsilon\cos\theta+\delta\sin\theta}$$

with $\theta=\theta(t)$ being subject to $\dot{\theta}=\lambda^2/r$, as above. Finally, we have

$$x^2+y^2 = (\varepsilon x+\delta y-c)^2$$

which is a degree 2 equation, and so the resulting trajectory is a conic.

PROOF. As already mentioned, this is a sort of "best of" the formulae found in the proof of Theorem 1.9. And in the hope of course that we have not forgotten anything. Finally, let us mention that the simplest illustration for this is the circular motion, and for details on this, not included in the above, we refer to the proof of Theorem 1.9. □

As a final question, we would like to understand how the various parameters appearing above, namely $\lambda, c, \varepsilon, \delta$, which via some basic math can only tell us more about the shape of the orbit, appear from the initial data. The formulae here are as follows:

THEOREM 1.11. *In the context of Theorem 1.10, and in polar coordinates, $x = r\cos\theta$, $y = r\sin\theta$, the initial data is as follows, with $R = r_0$:*

$$r_0 = \frac{c}{1+\varepsilon} \quad , \quad \theta_0 = 0$$

$$\dot{r}_0 = -\frac{\delta\sqrt{K}}{\sqrt{c}} \quad , \quad \dot{\theta}_0 = \frac{\sqrt{Kc}}{R^2}$$

$$\ddot{r}_0 = \frac{\varepsilon K}{R^2} \quad , \quad \ddot{\theta}_0 = \frac{4\delta K}{R^2}$$

The corresponding formulae for the affine coordinates x, y can be deduced from this. Also, the various motion parameters c, ε, δ and $\lambda = \sqrt{Kc}$ can be recovered from this data.

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

(1) As mentioned in Theorem 1.10, the object m begins its movement on Ox. Thus we have $\theta_0 = 0$, and from this we get the formula of r_0 in the statement.

(2) Regarding the initial speed now, the formula of $\dot{\theta}_0$ follows from:

$$\dot{\theta} = \frac{\lambda}{r^2} = \frac{\sqrt{Kc}}{r^2}$$

Also, in what concerns the radial speed, the formula of $\dot{r}_0$ follows from:

$$\begin{aligned}\dot{r} &= \frac{c(\varepsilon\sin\theta - \delta\cos\theta)\dot{\theta}}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2}\\ &= \frac{c(\varepsilon\sin\theta - \delta\cos\theta)}{c^2/r^2}\cdot\frac{\sqrt{Kc}}{r^2}\\ &= \frac{\sqrt{K}(\varepsilon\sin\theta - \delta\cos\theta)}{\sqrt{c}}\end{aligned}$$

(3) Regarding now the initial acceleration, by using $\dot{\theta} = \sqrt{Kc}/r^2$ we find:

$$\ddot{\theta} = -2\sqrt{Kc}\cdot\frac{2r\dot{r}}{r^3} = -\frac{4\sqrt{Kc}\cdot\dot{r}}{r^2}$$

In particular at $t=0$ we obtain the formula in the statement, namely:

$$\ddot{\theta}_0=-\frac{4\sqrt{Kc}\cdot\dot{r}_0}{R^2}=\frac{4\sqrt{Kc}}{R^2}\cdot\frac{\delta\sqrt{K}}{\sqrt{c}}=\frac{4\delta K}{R^2}$$

(4) Also regarding acceleration, with $\lambda=\sqrt{Kc}$ our main motion formula reads:

$$\ddot{r}=\frac{Kc}{r^2}\left(\frac{1}{r}-\frac{1}{c}\right)$$

In particular at $t=0$ we obtain the formula in the statement, namely:

$$\ddot{r}_0=\frac{Kc}{R^2}\left(\frac{1}{R}-\frac{1}{c}\right)=\frac{Kc}{R^2}\cdot\frac{\varepsilon}{c}=\frac{\varepsilon K}{R^2}$$

(5) Finally, the last assertion is clear, and since the formulae look better anyway in polar coordinates than in affine coordinates, we will not get into details here. □

All this is very nice, and congratulations of course, you know now more classical mechanics than the average nerd. Also, as a practical conclusion to all this, in case you ever forget the proof, but need to pull it out to someone, for whatever reasons, everything can be done with some math, by assuming that one of the masses is fixed, at 0.

1c. Inertial frames

Getting now back to Fact 1.2, we have to understand the notion of inertial frame appearing there. This is something quite tricky, as follows:

Definition 1.12. *An inertial frame is a frame where all the basic formulae, namely*

$$||F||=\frac{GM_1M_2}{||x_1-x_2||^2}\quad,\quad F=Ma\quad,\quad a=\dot{v}\quad,\quad v=\dot{x}\quad,\quad F_{12}=-F_{21}$$

hold, with the last formula standing for Newton's action-reaction principle.

To be more precise, the first 4 formulae are something very familiar, that we already heavily used, in the proof of Theorem 1.9. As for the last formula, also called Newton's third law, this expresses the fact that when an object 1 acts on an object 2, say via gravity, with force F_{12}, then object 2 acts as well on object 1, with force as follows:

$$F_{21}=-F_{12}$$

Regarding this third law of Newton, you probably heard of it since high school, but thinking a bit, this law does not really hold in the context of the computations that we did in the proof of Theorem 1.9. Indeed, assuming as there that M_1 is fixed at 0, its acceleration is $\ddot{0}=0$, and so the force acting on it is 0 too, which is certainly not the inverse of the force acting on M_2, which does exist, and drags M_2 on a conic.

All this might seem a bit perplexing, so let us formulate, as conclusion:

CONCLUSION 1.13. *Newton's third law is something quite tricky, and it's not about whether this law holds or not, but rather about if our frame is inertial or not.*

In short, you got the point, we are now into advanced physics, with no less than 5 useful formulae in our pocket, those in Definition 1.12. If these formulae mix well, we call the frame inertial, and this is very good. And if they don't, that is because the Newton third law is not satisfied in our frame, and that is good too, and we call our frame non-inertial, and we can of course still use it for computations.

But all this is perhaps a bit too abstract and mathematical. As a complement to this, here is an alternative definition for the inertial frames:

DEFINITION 1.14. *Your frame is inertial is your coffee does not spill.*

As examples of non-inertial frames, we have for instance elevators, rollercoaster rides, or you having a cat on your lap. As for inertial frames, you would probably say the Earth, but we will discuss this later, with the finding that, scientifically, if you fill your mug really up to the very top, it will spill, and so no good. But more on this later.

Getting back to math now, and to our inertial frames as axiomatized in Definition 1.12, here is a list of basic facts known about the inertial frames:

THEOREM 1.15. *The following hold, regarding the inertial frames:*

(1) *In the context of the 2-body problem, the standard frames used for computations, with M_1 or M_2 fixed, are not inertial.*
(2) *In fact, the linear combination frames or type $\lambda_1 M_1 + \lambda_2 M_2$, including the center of mass frame, are all non-inertial.*

PROOF. These facts are all well-known, the idea being as follows:

(1) This was something already discussed above, but let us present now the full mathematical proof. We want to check whether the forces between M_1, M_2 satisfy:

$$F_{12} = -F_{21} = \frac{GM_1M_2(x_1 - x_2)}{||x_1 - x_2||^3}$$

In the case of the frame centered at M_1, the formula $F_{12} = -F_{21}$ certainly does not hold, because the acceleration of M_1 is in this case $\ddot{0} = 0$, and so no force acting upon it, at least from our calculus viewpoint. The same holds for the frame centered at M_2.

(2) In the general case now, with parameters λ_1, λ_2 satisfying $\lambda_1 + \lambda_2 = 0$, as in the statement, the positions of our bodies M_1, M_2 are:

$$z_1 = -\lambda_1 x \quad , \quad z_2 = \lambda_2 x$$

Thus the forces acting upon M_1, M_2, computed according to calculus, are:

$$F_{21} = -M_1 \lambda_1 \ddot{x} \quad , \quad F_{12} = -M_2 \lambda_2 \ddot{x}$$

Thus, in order to have $F_{12} = -F_{21}$, the parameters λ_1, λ_2 must satisfy:

$$M_1\lambda_1 = M_2\lambda_2$$

But these are exactly the parameters of the center of mass of the system formed by M_1, M_2. But the center of mass frame is not inertial either, because due to the fact that we performed a dilation, the magnitude of $F_{12} = -F_{21}$ is not the correct one. $\square$

Some further interesting aspects of the inertial frames appear in connection with the angular momentum. In order to discuss this, we will need some math, as follows:

DEFINITION 1.16. *The vector product of two vectors in* $\mathbb{R}^3$ *is given by*

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

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

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

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

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

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

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

$$||x \times y|| \sim ||x||, ||y||$$

$$x \times x = 0$$

$$e_1 \times e_2 = e_3$$

So, here's how it works. We're looking for a vector $x \times y$ whose length is proportional to those of x, y. But now the second formula tells us that the angle θ between x, y must be involved via $0 \to 0$, and so the factor can only be $\sin\theta$. And with this we're almost there, it's just a matter of choosing the orientation, and this comes either from the right-hand rule (or perhaps left-hand rule, do I remember right?) or from $e_1 \times e_2 = e_3$.

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

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

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

The interest in the momentum comes from the following result:

THEOREM 1.17. *In the gravitational* 2*-body problem, the angular momentum*

$$J = x \times p$$

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

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

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

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

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

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

This was for the basic theory of the angular momentum, and for more, you can prove for instance, as an exercise, that Theorem 1.17 leads to the Kepler 2 and Kepler 3 laws, as claimed in Theorem 1.9. Now back to our inertial frames, all this is interesting for us, because we can formulate, as a completement to Theorem 1.15:

THEOREM 1.18. *The following hold, regarding the inertial frames:*

1. *The total angular momentum of a system of bodies $M_1, \ldots, M_k$ is conserved, when computed with respect to an inertial frame.*
2. *However, this is not the end of the story with angular momentum, because at $k = 2$, this is conserved as well in the frames of type $\lambda_1 M_1 + \lambda_2 M_2$.*

PROOF. As before with Theorem 1.15, all this is elementary, as follows:

(1) Our inertial frame assumption tells us that we can use at will all formulae in Definition 1.12, and by using them, and notably using $F_{12} = -F_{21}$, we obtain:

$$\begin{aligned}\dot{J} &= \sum_i x_i \times \sum_{j\neq i} F_{ji}\\ &= \sum_{i<j} x_i \times F_{ji} + x_j \times F_{ij}\\ &= \sum_{i<j} x_i \times F_{ji} - x_j \times F_{ji}\\ &= \sum_{i<j} (x_i - x_j) \times F_{ji}\\ &= 0\end{aligned}$$

Now since we have $\dot{J} = 0$, the angular momentum J is conserved, as claimed.

(2) This follows from a similar computation, and more specifically from the computation from the proof of Theorem 1.17, which gives $\dot{J} = 0$, and so conservation too. □

All in all, the conclusion would be that inertial frames are a rare asset, and also that the conservation of angular momentum, although not proving that the frame is inertial, is a good test for inertiality. Quite modest all this, but not bad, as a starting point.

So, let us start hunting for inertial frames. A simple idea here is to start with a familiar problem, such as the 2-body one, written in some standard non-inertial frame that is well-adapted for computations, and then try to transform that non-inertial frame into an inertial one. And regarding now the transformations, we can use here:

PROPOSITION 1.19. *Any frame can be obtained from another frame, by performing the following operations, which are independent of each other:*

(1) *Changing the origin,* $x \to x - d$.
(2) *Rescaling the coordinates,* $x \to r \cdot x$.
(3) *Rotating the coordinates,* $x \to Ux$ *with* $U \in O_3$.

PROOF. This is something obvious, written perhaps in a somewhat fancy way:

(1) This amounts in moving the origin 0 at an arbitrary location $d \in \mathbb{R}^3$.

(2) This amounts in rescaling the coordinates by $r_1, r_2, r_3 \in \mathbb{R}$, respectively.

(3) This amounts in rotating the frame, O_3 being the group of 3D rotations.

But these three operations clearly allow passing from one frame to another, as claimed, and we are led to the conclusion in the statement. □

Let us first examine the 2-body problem. Here we know that the center of mass frame, obtained via a transformation (1), is the best one that we have, so the problem is whether we can make it inertial using (2). And here, we have the following result:

THEOREM 1.20. *In the standard context of the 2-body problem, by moving the origin at the center of mass, and then rescaling all three coordinates by $r = r(t) \in \mathbb{R}$ given by*

$$(rx)'' = -\frac{GM_1(M_1 + M_2)(rx)}{||(rx)^3||}$$

with this meaning that $y = rx$ must be the trajectory of a body of mass M_1 travelling around a body of mass $M_1 + M_2$, the resulting frame is inertial.

PROOF. We will be actually looking at all the inertial frames that can be obtained via the operations (1) and (2) in Proposition 1.19, and only at the end we will particularize to the frame discussed in the statement. Our study goes as follows:

(1) Assume that M_1 is fixed at 0, and that M_2 moves around it, with position vector $x \in \mathbb{R}^3$. By changing the origin at $d \in \mathbb{R}^3$, then rescaling the coordinates by $r \in \mathbb{R}^3$, as in Proposition 1.19 (1,2), the new coordinates of M_1, M_2 are as follows:

$$z_1 = -r \cdot d \quad , \quad , z_2 = r \cdot (x - d)$$

We want to check whether the forces between M_1, M_2 satisfy the following formula:

$$M_1 \ddot{z}_1 = -M_2 \ddot{z}_2 = \frac{GM_1M_2(z_2 - z_1)}{||z_2 - z_1||^3}$$

By using the above formulae of z_1, z_2, this formula to be checked reads:

$$-M_1(r \cdot d)'' = -M_2(r \cdot x)'' + M_2(r \cdot d)'' = \frac{GM_1M_2(r \cdot x)}{||r \cdot x||^3}$$

(2) This looks complicated, so let us look for a uniform solution, $r = (r, r, r)$. In this case the componentwise dot product is a usual product, and our equation becomes:

$$-M_1(rd)'' = -M_2(rx)'' + M_2(rd)'' = \frac{GM_1M_2x}{r^2||x||^3}$$

By using now the formula of gravity in the initial frame, this equation becomes:

$$-M_1(rd)'' = -M_2(rx)'' + M_2(rd)'' = -M_1 \cdot \frac{x''}{r^2}$$

But this formula can be written, more conveniently, as follows:

$$(rd)'' = \frac{M_2}{M_1 + M_2}(rx)'' = \frac{x''}{r^2}$$

Thus, we have our equations, which look good, and we can now solve for r on the right, and then solve for d on the left, as to get our inertial frame.

(3) Let us first solve for r on the right. It is convenient here to replace x'' by the Newtonian gravitation formula it came from, and our equation becomes:

$$\frac{M_2}{M_1+M_2}(rx)'' = -\frac{GM_1M_2x}{r^2||x||^3}$$

But this equation can be written in the following more convenient way:

$$(rx)'' = -\frac{GM_1(M_1+M_2)(rx)}{||(rx)^3||}$$

We conclude that $y = rx$ must be the trajectory of a body of mass M_1 travelling around a body of mass $M_1 + M_2$, with arbitrary initial data y_0, w_0.

(4) Let us solve now for d on the left, in the equations found in (2) above. Here the situation is very simple, the solutions rd being as follows, with $a, b \in \mathbb{R}^3$:

$$rd = \frac{M_2}{M_1+M_2}rx + at + b$$

Thus, with r being as in (3) above, the solutions are as follows, with $a, b \in \mathbb{R}^3$:

$$d = \frac{M_2}{M_1+M_2}x + \frac{at+b}{r}$$

Thus with c being the center of mass, the formula is as follows, with $a, b \in \mathbb{R}^3$:

$$d = c + \frac{at+b}{r}$$

Now by taking $a = b = 0$, we are led to the conclusion in the statement. □

As a comment now, the above computation was certainly successful, but its technical details raise grim perspectives about the k-body case, with $k \geq 3$. Indeed, the equations of motion will be far more complicated, and we will not be able to perform even the simplest algebraic manipulations done in the above. As a second issue, we cannot hope for a simple answer, via a uniform rescaling $r = (r, r, \ldots, r)$, simply because the parameter $M_1(M_1 + M_2)$ appearing in Theorem 1.20 has no k-analogue, when $k \geq 3$. And finally, above everything, at $k \geq 3$ we have no solution to the problem that we can rely upon.

Regarding now the conservation of the angular momentum, we have here:

THEOREM 1.21. *In the context of the 2-body problem, by moving the origin at $d \in \mathbb{R}^3$, then uniformly rescaling by $r \in \mathbb{R}$, the angular momentum is conserved when:*

$$M_2\left(x \times (rd)'' + d \times (rx)'' - x \times (rx)''\right) = (M_1+M_2) \cdot d \times (rd)''$$

In the particular case where $d = \lambda x$, with λ being constant, these equations reduce to

$$(rx)'' \sim x$$

and r constant, as well as the inertial frame in Theorem 1.20, provide solutions.

PROOF. As before with the study in Theorem 1.20, we will do things slowly, by looking for explicit solutions only at the end. Our study goes as follows:

(1) Assume that M_1 is fixed at 0, and that M_2 moves around it, with position vector $x \in \mathbb{R}^3$. By changing the origin at $d \in \mathbb{R}^3$, then uniformly rescaling the coordinates by $r \in \mathbb{R}$, the new coordinates of M_1, M_2 are as follows:

$$z_1 = -rd \quad , \quad , z_2 = r(x - d)$$

Thus, the derivative of the total angular momentum is given by:

$$\begin{aligned} J' &= J_1' + J_2' \\ &= M_1 \cdot (-rd) \times (-(rd)'') + M_2 \cdot (rx - rd) \times ((rx)'' - (rd)'') \\ &= r(M_1 + M_2) \cdot d \times (rd)'' - rM_2\,(x \times (rd)'' + d \times (rx)'' - x \times (rx)'') \end{aligned}$$

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

(2) In the particular case $d = \lambda x$, with λ constant, the above formula of J' reads:

$$\begin{aligned} J' &= r(M_1 + M_2)\lambda^2 \cdot x \times (rx)'' - rM_2\,(2\lambda x \times (rx)'' - x \times (rx)'') \\ &= r\left((M_1 + M_2)\lambda^2 + M_2(1 - 2\lambda)\right) \cdot x \times (rx)'' \\ &= r\left((M_1 + M_2)\lambda^2 - 2M_2\lambda + M_2\right) \cdot x \times (rx)'' \end{aligned}$$

The discriminant on the left being $\Delta = -4M_1M_2 < 0$, our only hope is $x \times (rx)'' = 0$, which amounts in saying that we must have $(rx)'' \sim x$, as stated.

(3) Finally, in what regards the explicit solutions mentioned at the end, we already know from the above that the conservation of J holds for them. But these are now all particular cases of the present result. Indeed, r constant gives $(rx)'' \sim x$, and at $r = 1$ we obtain the frames of type $\lambda_1 M_1 + \lambda_2 M_2$, studied before. Also, with $d = c$ being the center of mass, and with r being as in Theorem 1.20, we have $(rx)'' \sim rx \sim x$, and the frame here coincides with the inertial one constructed in Theorem 1.20, as desired. □

As before with Theorem 1.20, the above computation was something successful in the 2-body case, unifying things that we knew so far. But also as before with Theorem 1.20, the details of the proof raise grim perspectives on what will happen for k bodies, with $k \geq 3$. Again, too many forces acting there, things too complex, and so on.

Before leaving the subject, let us formulate as well:

THEOREM 1.22. *Given bodies $M_1, \ldots, M_k$ moving in an inertial frame, with coordinates $x_1, \ldots, x_k$, their total angular momentum, which is constant, is given by*

$$J = \sum_i M_i \cdot c \times \dot{c} + \sum_i M_i \cdot y_i \times \dot{y}_i$$

where $y_i = x_i - c$ are the relative coordinates with respect to the center of mass c. Thus J is the momentum of the average system, plus the momentum computed at c.

PROOF. The fact that J is constant is something that we know, coming from:

$$\begin{aligned}\dot{J} &= \sum_{i<j} x_i \times F_{ji} + x_j \times F_{ij}\\ &= \sum_{i<j}(x_i - x_j) \times F_{ji}\\ &= 0\end{aligned}$$

Regarding now the second assertion, consider the center of mass c, and write the coordinates of $M_1, \ldots, M_k$ in the form $x_i = c + y_i$. We have then:

$$\begin{aligned}J &= \sum_i M_i \cdot x_i \times \dot{x}_i\\ &= \sum_i M_i \cdot (c + y_i) \times (\dot{c} + \dot{y}_i)\\ &= \sum_i M_i \cdot c \times \dot{c} + \sum_i M_i y_i \times \dot{c} + c \times \sum_i M_i \dot{y}_i + \sum_i M_i \cdot y_i \times \dot{y}_i\end{aligned}$$

Now since c is the center of mass we have $\sum_i M_i y_i = 0$, and we are left with:

$$J = \sum_i M_i \cdot c \times \dot{c} + \sum_i M_i \cdot y_i \times \dot{y}_i$$

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

1d. Rotating frames

As a last general topic regarding frames, let us examine now the rotations of frames, and what can be done with them, a subject that we have not got into, so far. With the remark that this question is of particular interest for us humans, living on Earth, which rotates. There are many things that can be said here, and we first have:

THEOREM 1.23. *Assume that a 3D body rotates along an axis, with angular speed w. For a fixed point of the body, with position vector x, the usual 3D speed is*

$$v = \omega \times x$$

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

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

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

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

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

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

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

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

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

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

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

THEOREM 1.24. *Assuming as before that a 3D body rotates along an axis, the acceleration of a moving point on the body, computed by O as before, is given by*

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

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

PROOF. This comes by using twice the formulae in Theorem 1.23, as follows:

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

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

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

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

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

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

THEOREM 1.25. *The acceleration of an object m subject to a force F is given by*

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

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

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

We refer to any standard undergraduate mechanics book, such as Feynman [**32**], Kibble [**54**] or Taylor [**88**] for more on the above, including various numerics on what happens here on Earth, the Foucault pendulum, history of all this, and many other things. Let us just mention here, as a basic illustration for all this, that a rock dropped from 100m deviates about 1cm from its intended target, due to the formula in Theorem 1.25.

All this is quite interesting, but it's getting late, time to stop our hunting for inertial frames, get back home, and grill a steak bought at the supermarket. The day has been long, and our first discovery was that the inertial frames do not really exist, mathematically speaking. That is, given a concrete math question, formulated in a non-inertial frame, you cannot really prove via math that an inertial frame for it exists.

As for our second discovery, this was the fact that the inertial frames do not really exist in the real life either. There is rotation everywhere, and a myriad other problems, and there might be an inertial frame suspended somewhere, out there in the universe, in the void, but we will certainly not be able to find it, with our technology.

This being said, the notion of inertial frame perfectly makes sense, abstractly, and we will use this notion, as in Fact 1.2 (2), in order to talk about the speed of light c.

1e. Exercises

Exercises.

CHAPTER 2

Speed addition

2a. Einstein summation

We have seen in chapter 1 that classical mechanics needs a fix. To be more precise, as explained there, based on various experiments by Fizeau, then Michelson-Morley and others, and on some physics by Maxwell and Lorentz, which are actually still to be discussed, and on some thinking too, Einstein came upon the following principles:

Fact 2.1 (Einstein principles). *The following happen:*

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

But these principles show, via a simple Gedankenexperiment involving a light bulb being switched on inside a moving train, explained in chapter 1, that we formally have $c + v = c$, for any speed $v > 0$, which is in contradiction with classical mechanics, and even with common sense, and with mathematics itself.

Obviously, in view of this, there are many things to be done. Since speed is distance over time, $v = d/t$, maybe something is wrong with distance d, or with time t, or with both. Or maybe something is wrong with the formula $v = d/t$ itself. And so on. In addition, even when assuming that we manage to fix all this, speeds, then comes a review of the momentum p, and of energy E, which crucially depend on speed. And finally, even with all this done, we will still have to see if this new theory that we are building, let's call it "special relativity", is compatible with gravity, and electromagnetism.

All this looks overly complicated, so many things to be done, and not even clear how to start, and what strategy to follow. So, time to ask the cat. And cat says:

Cat 2.2. *Math is trivial, physics is complicated. So invent some math for*

$$c + v = c$$

I bet you can do that, and check later with physics if your solution is correct.

As usual with what cat says, this sounds very wise. Cat surely knows some physics, like all felines and other predators do, but for slow and peaceful people like us, humans, bears and so on, doing some mathematics first is recommended.

So, nice idea, and we will do this. Following what cat says, and also Einstein's book [28], which seems written under cat infuence, with a lot of modest mathematics and thinking inside, we will develop our theory in two steps, as follows:

(1) The first step, called special relativity, consists in forgetting gravity, and fixing the basic laws of motion, as to have $c+v=c$. This will be actually not that hard, with just a little bit of algebra and geometry involved, and we will explain this, along with some consequences, and with due physics verifications, in Parts I-II of this book.

(2) As for the question coming afterwards, that of including gravity into this new theory of motion, this is something more complicated, called general relativity. We will discuss this, again along with some applications, for the most to celestial mechanics, where general relativity can be observed in practice, in Parts III-IV of this book.

Getting started now, $c+v=c$ is obviously about adding speeds, so this topic, adding speeds, will be the first one that we will discuss. In the classical case, we have:

PROPOSITION 2.3. *The classical speeds add according to the Galileo formula*

$$v_{AC}=v_{AB}+v_{BC}$$

where v_{AB} denotes the relative speed of A with respect to B.

PROOF. This is clear indeed from the definition of speed, and very intuitive. □

In order to find the fix, we will first discuss the 1D case, and leave the 3D case, which is a bit more complicated, for later. We will use two tricks. First, let us forget about absolute speeds, with respect to a given frame, and talk about relative speeds only. In this case we are allowed to sum only quantities of type v_{AB},v_{BC}, and we denote by $v_{AB}+_g v_{BC}$ the corresponding sum v_{AC}. With this convention, the Galileo formula becomes:

$$u+_g v=u+v$$

As a second trick now, observe that this Galileo formula holds in any system of units. In order now to deal with our problems, basically involving high speeds, it is convenient to change the system of units, as to have $c=1$. With this convention our $c+v=c$ problem becomes $1+v=1$, and the solution to it is quite obvious, as follows:

THEOREM 2.4. *If we define the Einstein sum $+_e$ of relative speeds by*

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

in $c=1$ units, then we have the formula $1+_e v=1$, valid for any v.

PROOF. This is obvious indeed from our definition of $+_e$, because if we plug in $u = 1$ in the above formula, we obtain as result:

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

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

Summarizing, we have solved our problem. In order now to formulate a final result, we must do some reverse engineering, by waiving the above two tricks. First, by getting back to usual units, $v \to v/c$, our new addition formula becomes:

$$\frac{u}{c} +_e \frac{v}{c} = \frac{\frac{u}{c} + \frac{v}{c}}{1 + \frac{u}{c} \cdot \frac{v}{c}}$$

By multiplying by c, we can write this formula in a better way, as follows:

$$u +_e v = \frac{u+v}{1 + uv/c^2}$$

In order now to finish, it remains to get back to absolute speeds, as in Proposition 2.3. And by doing so, we are led to the following result:

THEOREM 2.5. *If we sum the speeds according to the Einstein formula*

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

then the Galileo formula still holds, approximately, for low speeds

$$v_{AC} \simeq v_{AB} + v_{BC}$$

and if we have $v_{AB} = c$ or $v_{BC} = c$, the resulting sum is $v_{AC} = c$.

PROOF. We have two assertions here, which are both clear, as follows:

(1) Regarding the first assertion, if we are at low speeds, $v_{AB}, v_{BC} << c$, the correction term $v_{AB}v_{BC}/c^2$ dissapears, and we are left with the Galileo formula, as claimed:

$$\begin{aligned} v_{AC} &= \frac{v_{AB} + v_{BC}}{1 + v_{AB}v_{BC}/c^2} \\ &\simeq \frac{v_{AB} + v_{BC}}{1 + 0} \\ &= v_{AB} + v_{BC} \end{aligned}$$

(2) As for the second assertion, this follows from the above discussion, but let us doublecheck this, to make sure that we have made no mistake. With $v_{AB} = c$ we get:

$$v_{AC} = \frac{c + v_{BC}}{1 + v_{BC}/c} = c$$

Similarly, assuming $v_{BC}=c$, the resulting speed that we get is:

$$v_{AC}=\frac{v_{AB}+c}{1+v_{AC}/c}=c$$

Thus, no mistakes in our mathematics, and we are done. □

As a conclusion, we have solved the problem, mathematically. All this is very nice, and fits with the Einstein principles from Fact 2.1, and to be more precise, fixes the laws of motion, as to make them fit with that principles. But the problem is now, does all this math really correspond to physics? And the answer here is fortunately yes, due to:

Fact 2.6. *The Einstein addition formula is correct.*

This is of course a physics fact, based on several things, which are worth a detailed discussion, as follows:

(1) First we have the Galileo formula, thought before to be exact, $v_{AC}=v_{AB}+v_{BC}$, and some numerics were to be done here, in order to convince the audience that this formula might well be $v_{AC}\simeq v_{AB}+v_{BC}$, due to the inevitable tiny errors in our speed measuring machinery, and with the correction term which might well be the Einstein one, $v_{AB}v_{BC}/c^2$ as above. This of course does not prove the Einstein formula, but shows at least that it is compatible with the Galileo formula.

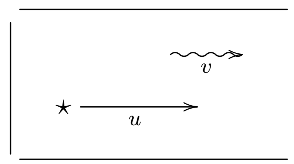
(2) As a second piece of evidence now, which is considerably more serious, the Einstein formula is compatible with Fact 2.1, which is a physics fact. Thus, we are here at the point where the Einstein formula is a "beautiful mathematical fix, compatible with everything known before, namely the Einstein principles and the Galileo formula".

(3) As a third piece of evidence, which is however something more advanced, the Einstein addition formula, or rather the concepts of time dilation and length contraction that will emerge from it, as we will soon discover in chapter 3, are compatible with the basic physics of electromagnetism, as developed by Maxwell, and Lorentz.

(4) Finally, as a fourth and ultimate piece of evidence, various observations and experiments at very high speeds, $v_{AC}\simeq c$, which needless to say, were as usual subject to some tolerances in our speed measuring machinery, confirmed the Einstein formula.

So this was for the story, a bit oversimplified because the real story involved in fact the 3D generalization of Theorem 2.5 and of Fact 2.6, that we have chosen here in this book to discuss later. This being said, as something very concrete in 1D, in relation with (4), which according to Einstein himself [28] was the chief inspiration and verification of his relativity theory, we have the following key experiment of Fizeau:

EXPERIMENT 2.7 (Fizeau, 1851). *Assume that light moves through a liquid at speed $u < c$. Then, when this liquid moves through a tube at speed $v > 0$,*



the observed speed of light is not the Galilean $u +_g v = u + v$, but rather

$$u +_f v = u + v\left(1 - \frac{1}{n^2}\right)$$

where $n = c/u$ is the index of refraction of the liquid.

You must agree with me that this looks very good, especially in what regards the first part, with the observed speed by Fizeau being not the Galilean one, $u +_f v \neq u +_g v$, but then with the second part too, with Fizeau's sum $u +_f v$ looking quite similar to the Einstein sum $u +_e v$. So, let us do now the math, and compare what Fizeau and Einstein say. The result here, which is certainly a success, if you are a bit familiar with the difficulties of experimental physics, say via daily cooking at home, is as follows:

THEOREM 2.8. *The Fizeau speed summation, which in $c = 1$ units is*

$$u +_f v = u + v - u^2 v \simeq (u + v)(1 - uv)$$

is compatible with the Einstein speed summation, which in $c = 1$ units is

$$u +_e v = \frac{u + v}{1 + uv} \simeq (u + v)(1 - uv)$$

with the approximations coming from $u >> v$, and from $1/(1 + x) \simeq 1 - x$.

PROOF. This is something rather self-explanatory, but let us work out the details. In $c = 1$ units the index of refraction of the liquid is $n = 1/u$, and we have:

$$\begin{aligned} u +_f v &= u + v\left(1 - \frac{1}{(1/u)^2}\right) \\ &= u + v(1 - u^2) \\ &= u + v - u^2 v \\ &\simeq u + v - u^2 v - uv^2 \\ &= (u + v)(1 - uv) \end{aligned}$$

To be more precise here, we have used, for the approximation at the end:

$$v << u \implies uv^2 << u, v, u^2 v$$

As for the processing of the Einstein formula, this simply uses $1/(1 + x) \simeq 1 - x$. $\square$

We will be back to this, and to other experiments supporting the Einstein summation formula, in 1D and in 2D, 3D as well, on several occasions, throughout this book.

Getting back now to the Einstein summation formula from Theorem 2.5, or rather to its more compact formulation from Theorem 2.4, that formula, while looking very simple, is in fact quite subtle, and must be handled with care. Indeed, we have:

THEOREM 2.9. *The Einstein speed summation, written in $c = 1$ units as*

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

has the following properties:

(1) $u, v < 1$ *implies* $u + v < 1$.
(2) $\lambda u +_e \lambda v = \lambda(u + v)$ *fails.*
(3) $(u +_e v) +_e w = u +_e (v +_e w)$ *fails too.*

PROOF. All these assertions are elementary, as follows:

(1) This follows from the following formula, valid for any speeds u, v:

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

(2) This is clear, with the remark however that it works at $\lambda = -1, 0, 1$.

(3) This follows indeed from the following computation:

$$\begin{aligned} (u +_e v) +_e w &= \frac{u + v + w + uv}{1 + uv + uw + vw} \\ &\neq \frac{u + v + w + vw}{1 + uv + uw + vw} \\ &= u +_e (v +_e w) \end{aligned}$$

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

Summarizing, as said above, Einstein summation must be used with extreme care. Along the same lines, observe that the following formula holds in 1D:

$$u +_e v = v +_e u$$

However, as bad news, this formula fails in higher dimensions. More about this in a moment, when systematically investigating the 3D extension of the above.

2b. Three dimensions

All the above is very nice, but remember, takes place in 1D. So, time now to get seriously to work, and see what all this becomes in 3D. Expect of course a lot a vector calculus, as usual in relation with 3D problems, and in the hope that you love that.

As a main goal, we must review the Einstein speed summation formula:

QUESTION 2.10. *What is the correct analogue of the Einstein summation formula*

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

in 2 and 3 dimensions?

In order to discuss this question, let us attempt to construct $u +_e v$ in arbitrary dimensions, just by using our common sense and intuition, a bit as we did in 1D, with the advice Cat 2.2 in mind. As a first observation, when the vectors $u, v \in \mathbb{R}^N$ are proportional, we are basically in 1D, and so our addition formula must satisfy:

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

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

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

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

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

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

PUZZLE 2.11. *What vanishes when $u \sim v$, and then how to correctly define*

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

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

But the solution to the first question is well-known in 3D. Indeed, here we can use the vector product $u \times v$, that we met in chapter 1, which notoriously satisfies:

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

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

Puzzle 2.12. *How to correctly define the Einstein summation in* 3 *dimensions,*

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

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

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

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

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

Moving ahead now, as obvious task, we must "transport" the vector w to the plane spanned by u, v. But this is simplest done by taking the vector product with any vector in this plane, and so as a reasonable candidate for our correction term, we have:

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

Here $\alpha, \beta \in \mathbb{R}$ are some scalars to be determined, but let us take a break, and leave the computations for later. We did some good work, time to update our puzzle:

Puzzle 2.13. *How to define the Einstein summation in* 3 *dimensions,*

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

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

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

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

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

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

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

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

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

PROPOSITION 2.14. *The correction term* $\gamma_{uvw} = (\alpha u + \beta v) \times w$ *is given by*

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

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

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

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

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

Time now to get into the real thing, see what happens when $||u|| = 1$ and $||v|| = 1$, if we can get indeed $u +_e v = u$ and $u +_e v = v$. It is convenient here to do some reverse engineering. Regarding the first desired formula, namely $u +_e v = u$, we have:

$$\begin{aligned}
& u +_e v = u \\
\iff\ & u + v + \gamma_{uvw} = (1 + < u, v >)u \\
\iff\ & \gamma_{uvw} = < u, v > u - v \\
\iff\ & (\alpha < u, v > +\beta < v, v >)u - (\alpha < u, u > +\beta < u, v >)v = < u, v > u - v \\
\iff\ & \alpha = 1,\ \beta = 0,\ ||u|| = 1
\end{aligned}$$

Thus, with the parameter choice $\alpha = 1, \beta = 0$, we will have, as desired:

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

In what regards now the second desired formula, namely $u +_e v = v$, here the computation is almost identical, save for a sign switch, which after some thinking comes from our choice $w = u \times v$ instead of $w = v \times u$, clearly favoring u, as follows:

$$\begin{aligned}
& u +_e v = v \\
\iff\ & u + v + \gamma_{uvw} = (1 + < u, v >)v \\
\iff\ & \gamma_{uvw} = -u + < u, v > v \\
\iff\ & (\alpha < u, v > +\beta < v, v >)u - (\alpha < u, u > +\beta < u, v >)v = -u + < u, v > v \\
\iff\ & \alpha = 0,\ \beta = -1,\ ||v|| = 1
\end{aligned}$$

Thus, with the parameter choice $\alpha = 0, \beta = -1$, we will have, as desired:

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

All this is mixed news, because we managed to solve both our problems, at $||u|| = 1$ and at $||v|| = 1$, but our solutions are different. So, time to breathe, decide that we did enough interesting work for the day, and formulate our conclusion as follows:

Proposition 2.15. *When defining the Einstein speed summation in* 3D *as*

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

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

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

Proof. Here (1) and (2) follow from the above discussion, with the following choice for the correction term, by favoring the $||u|| = 1$ problem over the $||v|| = 1$ one:

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

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

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

As for (3) and (4), these are also clear from the above discussion, coming from the obvious lack of symmetry of our summation formula. □

What is next? Obviously, many things, including checking with mathematics and physics if our correction term, and final formula, are good, at least in some particular situations, and then especially, fixing the lack of symmetry in u, v of our formula. I don't know about you, but this dissymmetry thing is barely something I can live with.

Unfortunately, fixing the symmetry issue does not look easy at all, and you are of course free to do your computations too, in order to see that this does not look doable. Thus, we must acknowledge defeat, and ask the cat. And cat says:

Cat 2.16. *Why bothering about how zombies riding on elementary particles see you, better focus on the opposite problem, how you see these zombies.*

Which sounds very interesting, looks like we both missed the point, at high speeds $v \simeq c$ there might be indeed a dissymmetry between the observer and the observed. So, very nice, let's indeed be modest, and stay on planet Earth with our physics. We will be no longer looking for a symmetry fix, and the only job left will be that of checking, with mathematics and physics, if our formula in Proposition 2.15 is correct.

It is tempting here to go ahead with physics and checks, but remember from Cat 2.2 that as long as there is mathematics to do, we should not be lazy, and do that mathematics.

So, looking now at Proposition 2.15 from an abstract, mathematical perspective, there are still many things missing from there, which can be summarized as follows:

QUESTION 2.17. *Can we fine-tune the Einstein speed summation in* 3D *into*

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

with $\lambda \in \mathbb{R}$*, chosen such that* $||u|| = 1 \implies \lambda = 1$*, as to have:*

(1) $||u||, ||v|| < 1 \implies ||u +_e v|| < 1$.
(2) $||v|| = 1 \implies ||u +_e v|| = 1$.

All this is quite tricky, and deserves some explanations. First, if we add a scalar $\lambda \in \mathbb{R}$ into our formula, as above, we will still have, exactly as before:

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

On the other hand, we already know from our previous computations, those preceding Proposition 2.15, that if we ask for $\lambda \in \mathbb{R}$ to be a plain constant, not depending on u, v, then $\lambda = 1$ is the only good choice, making the following formula happen:

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

But, and here comes our point, $\lambda = 1$ is not an ideal choice either, because it would be nice to have the properties (1,2) in the statement, and these properties have no reason to be valid for $\lambda = 1$, as you can check for instance by yourself by doing some computations. Thus, the solution to our problem most likely involves a scalar $\lambda \in \mathbb{R}$ depending on u, v, and satisfying the following condition, as to still have $||u|| = 1 \implies u +_e v = u$:

$$||u|| = 1 \implies \lambda = 1$$

Looks like we have a very good idea here, so time to say hello to kit-kat, and share with him our latest discoveries. However, cat, unfazed, declares:

CAT 2.18. *Gauge invariance gives you everything. By the way, you'll also understand why things do not commute.*

Well, that was a nice try, but cat has spoken, and this was the type of advanced comment I was afraid of. The point indeed is that, thinking well, our construction and study so far of $u +_e v \in \mathbb{R}^3$ ignore one important thing, namely the action of the various rotations $U \in O_3$ on the vectors $u, v \in \mathbb{R}^3$, and on the resulting vector $u +_e v \in \mathbb{R}^3$. And so, as cat says, if we decide that we want our theory of $u +_e v$ to behave well with respect to rotations, we will probably end up with a goldmine of useful requirements, formulae and results, most likely solving all our past, present and future problems.

This being said, rotations in $\mathbb{R}^3$, and sometimes even in $\mathbb{R}^2$, can be a quite complicated business, and since I have this feeling that we are very close, with our elementary way of doing things, we will just keep doing that, and leave what cat says for later.

So, getting back to our problem, we must construct $\lambda \in \mathbb{R}$ satisfying:

$$||u|| = 1 \implies \lambda = 1$$

Obviously, as simplest answer, λ must be some well-chosen function of $||u||$, or rather of $||u||^2$, because it is always better to use square norms, when possible. But then, with this idea in mind, after a few computations we are led to the following solution:

$$\lambda = \frac{1}{1 + \sqrt{1 - ||u||^2}}$$

Summarizing, final correction done, and with this being the end of mathematics, we did a nice job, and we can now formulate our findings as a theorem, as follows:

THEOREM 2.19. *When defining the Einstein speed summation in* 3D *as*

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

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

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

PROOF. This follows from the above discussion, as follows:

(1) This is something that we know from Proposition 2.15, coming from:

$$u \sim v \implies u \times v = 0 \implies u +_e v = \frac{u + v}{1 + < u, v >}$$

(2) This is something which follows from some computations, and we will be back to this, with details, in a moment, directly in arbitrary N dimensions. However, let us briefly present these computations. Let us set, as in the statement:

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

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

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

Taking now the squared norm gives the following formula:

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

By expanding the two squares and multiplying at the end, and then simplifying everything, we obtain a quite compact formula, as follows:

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

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

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

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

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

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

(3) This is something that we know from Proposition 2.15, coming from:

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

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

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

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

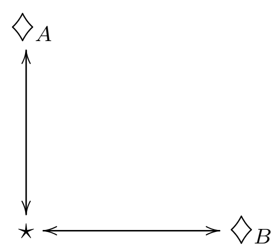
(6) This is again clear, from the obvious lack of symmetry of our formula. □

All this is very nice, and time to remember now that we were supposed to do physics, in this book. But, good news here, our formula is indeed the correct one:

FACT 2.20. *The formula in Theorem 2.19 is correct, physically speaking.*

There are of course many things that can be said here, and we will be back to this, on several occasions, but on the bottom line, the point is that the formula in Theorem 2.19 is indeed verified, by a certain number of experiments, that a full freight train payload of theory cannot replace. Of particular importance is the following experiment by Michelson and Morley, coming as a continuation of Experiment 2.7 of Fizeau, from 1851:

EXPERIMENT 2.21 (Michelson-Morley, 1887). *When sending beams of light to mirrors which are at the same distance $L > 0$, but are positioned differently*



there return after slightly different times, and with the mathematics of these return times confirming the Einstein summation formula in 2D, and so in 3D as well.

Here both the experiments, which came in different configurations based on the above, getting more and more complicated over the time, and the mathematics, involving the rotational speed of the Earth, are quite complicated, so you will have to trust me here. In addition, as an interesting historical detail, Michelson and Morley originally wanted to prove something else, but their experiment gradually evolved into what is described above, providing a solid verification for Einstein's future relativity theory.

As another comment, Albert Michelson received in 1907 the Nobel Prize in Physics, for his remarkable lifetime work on light, with Morley and many others, becoming in this way the first American to win the Nobel Prize in science. Which is very good, but unfortunately the enthusiasm for this went beyond the acceptable limits during the Cold War, with several US books about relativity forgetting to talk about Fizeau, and presenting the theory as "Michelson-Morley, followed by Einstein", which is wrong.

Although I haven't checked with USSR books, I am pretty much sure that according to them, relativity theory comes from the Soviets. Perhaps even from Lenin himself, who famously declared "communism is Soviet power plus electrification of the country", which makes it clear that electricity, light and related physics are a USSR business.

In any case, for the whole story, the best is to learn it from Einstein himself [28].

2c. Questions, answers

All the above is very nice, and you might even think that done with this, but after some examination of our formula, several questions appear. Here is a list of such questions:

Questions 2.22. *In relation with the Einstein speed summation in* 3D*:*

(1) *What is the formula of* $||u +_e v||$*?*
(2) *Why* $||u||, ||v|| < 1 \implies ||u +_e v|| < 1$*?*
(3) *Honestly, why does* $u +_e v = v +_e u$ *fail?*
(4) *Also, what about other dimensions,* $N \neq 3$*?*

Here (1) is obviously just some math, so very nice, and we will eat this raw, as appetizer. Regarding now (2), here we know $||u +_e v|| \leq 1$ from physics, that is, from Fact 2.20 as stated, but for that strict estimate, and also for a full proof without physics, that can only come from (1). Regarding now (3), this is someting quite mind-boggling, most likely requiring both tricky mathematics and tricky physics, perhaps some tricky chemistry for long-term cooking too, and we will leave it for later, as main course.

As for (4), this might look a bit anecdotical, suitable for dessert, but after some thinking, this is in fact a very good problem, that we can start with. Indeed, we would obviously like to have, for many practical matters, a unified $N = 1, 3$ formula. But then also, again with practical applications in mind, what is the best $N = 2$ formula, and can we actually have a unified $N = 1, 2, 3$ formula? And then, with mathematics teaching us that many formulae are actually simpler for $N \in \mathbb{N}$ than for $N = 1, 2, 3$, why not simply looking for a N-dimensional formula, with $N \in \mathbb{N}$, and then studying that formula, in relation with (1,2,3). So, this would be my proposal, filing (4) as vodka, suitable at all stages of the meal. And too bad for the dessert, we'll do like Stone Age men, none.

In practice now, the problem is that the vector product $\times$ exists only in 3D. But we already have in fact an answer to this issue, coming from the following formula:

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

So, based on this, let us formulate the following definition, which is something purely mathematical, but coming with serious physics content:

DEFINITION 2.23. *The relativistic speed summation in $N\in\mathbb N$ dimensions is*

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

as usual in $c=1$ units.

As a first observation, this is the good formula at $N=1$, where the numerator or the correction term vanishes, $u^2v-u^2v=0$. At $N=3$ this is the good formula too, as we know from the above. Finally, this is also the good formula at $N=2$, because we take vectors $u,v\in\mathbb R^2$ and we complete with 0 entries, and make the sum in 3D, this is the same as summing in 2D, and then completing with a 0 entry.

So, problem solved, and we are now ready for Questions 2.22 (1,2). We have here:

THEOREM 2.24. *We have the following formula, in $N\in\mathbb N$ dimensions,*

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

and in particular, we have $||u||,||v||<1\implies ||u+_e v||<1$.

PROOF. This follows by computation, the details being as follows:

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

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

(2) Taking now the squared norm gives the following formula:

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

(3) Expanding the various squares gives the following formula:

$$\begin{aligned}||u+_e v||^2 &= \frac{1}{(1+<u,v>)^2(1+\delta)^2}\times\\ &\Big[(1+\delta)^2||u||^2+<u,v>^2||u||^2+2(1+\delta)<u,v>||u||^2\\ &+(1+\delta)^2||v||^2+||u||^4||v||^2-2(1+\delta)||u||^2||v||^2\\ &+2(1+\delta)^2<u,v>+2(1+\delta)<u,v>^2\\ &-2(1+\delta)||u||^2<u,v>-2<u,v>^2||u||^2\Big]\end{aligned}$$

(4) Rearranging terms and simplifying gives the following formula:

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

(5) But this can be written in a more compact form, as follows:

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

(6) Time now to remember that $\delta=\sqrt{1-||u||^2}$. Thus, we have $\delta^2=1-||u||^2$, and with this in mind, it is convenient to expand the first $(1+\delta)^2$ factor. We obtain:

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

(7) By using now $\delta^2=1-||u||^2$, this gives the following formula:

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

(8) At this point, it looks better to replace the $2-||u||^2$ factor by $1+\delta^2$. We get:

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

(9) Which is nice, looks like the gods of calculus are with us. We obtain:

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

(10) Thus, we proved the formula in the statement. Regarding now the last assertion, this follows from this. Indeed, assuming $||u||,||v||<1$, we have:

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

Thus, the last assertion holds too, and we are done. □

As a conclusion now, our N-dimensional speed summation formula from Definition 2.23 works just fine, and in connection with our original Questions 2.22, the only problem left there is to understand, conceptually, why $u+_e v=v+_e u$ fails. This is something quite tricky, that we will discuss later. Among others, this will lead as well to a more conceptual explanation for the summation formula from Definition 2.23, and also, to be honest with you, now that we did all that beautiful computations together, for the formula from Theorem 2.24. All this will come from the magic of the Lorentz transformation.

But more on that later, we are not ready yet for such things, the idea being that the Lorentz transformation is an outer space creature mixing space and time. And in the hope that you agree with me, although we are on our way of understanding relativity theory, mixing space and time still looks like a crazy idea, at least with our knowledge so far. We will only talk about such crazy things later, when becoming crazy ourselves.

2d. Further formulae

Let us go back now to 2D and 3D, with the aim of establishing some further formulae here, based on the above, which can be useful for computations. We first recall the general formulae that we have, established above in arbitrary $N\in\mathbb{N}$ dimensions:

THEOREM 2.25. *The Einstein summation formula in N dimensions is*

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

and the squared norm of this resulting speed is given by

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

with everything being expressed, as usual, in $c = 1$ units.

PROOF. This is indeed a summary of the main results that we have, so far. □

As a first observation, as already mentioned on numerous occasions, in the case where the vectors are proportional, $u \sim v$, the correction term dissapears, and we are led to the usual 1D formulae. Along the same lines, some simplifications appear as well when the vectors are orthogonal, $u \perp v$, with the result here being as follows:

PROPOSITION 2.26. *When the vectors are orthogonal, $u \perp v$, we have*

$$u +_e v = u + v - \frac{||u||^2}{1 + \sqrt{1 - ||u||^2}} \cdot v$$

and the squared norm of this resulting speed is given by

$$||u +_e v||^2 = ||u + v||^2 - ||u||^2||v||^2$$

with everything being expressed, as usual, in $c = 1$ units.

PROOF. This follows indeed from Theorem 2.25, applied with $< u, v >= 0$. □

More interestingly now, in the general case, where u, v are not parallel or orthogonal, in 3D there is a simplification appearing in the last formula in Theorem 2.25, coming from the magic of the vector product, the final result on the subject being as follows:

THEOREM 2.27. *In 3 dimensions we have the norm formula*

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

valid for any two vectors $u, v \in \mathbb{R}^3$.

PROOF. This is something very nice, coming from our previous norm formula, from Theorem 2.25, and from Pythagoras. Let us recall indeed that the magnitude of an arbitrary vector product is as follows, with θ being the angle between the vectors:

$$||u \times v|| = ||u|| \cdot ||v|| \cdot |\sin\theta|$$

On the other hand, we have a similar formula for the scalar product, namely:

$$< u, v >= ||u|| \cdot ||v|| \cdot \cos\theta$$

Now by raising both the above quantities to the square, and summing, we obtain, by using the formula $\sin^2\theta+\cos^2\theta=1$, coming itself from Pythagoras:

$$||u\times v||^2+<u,v>^2=||u||^2||v||^2$$

Thus, the squared norm of the vector product is given by the following formula:

$$||u\times v||^2=||u||^2||v||^2-<u,v>^2$$

But this is exactly what we need, because our general formula for $||u+_e v||^2$ makes appear the quantity on the right. By replacing that with $||u\times v||^2$, we obtain the result. □

Moving ahead, an idea coming from Proposition 2.26 would be that of decomposing the problem into a $u\sim v$ part, and a $u\perp v$ part. However, this can be properly done in 2D only, so let us discuss now 2D. Regarding the formula of $u+_e v$ itself, there is no notable simplification when assuming that we are in 2D. However, in what regards the formula of $||u+_e v||^2$, there is some work to be done here, and we have:

Theorem 2.28. *In 2 dimensions the squared norm formula*

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

can be best applied by using the following identity, with $u=\binom{a}{b}$, $v=\binom{c}{d}$,

$$||u||^2||v||^2-<u,v>^2=(ad-bc)^2$$

with the remark that this further simplifies when one of a,b,c,d *is zero.*

Proof. Here the formula in the statement is clear, and you can also recover this by passing to 3D, via Theorem 2.27, if you really want to. As for the last remark, which is of course something trivial, this is of both theoretical and practical importance, because up to a rotation in the plane, we can always assume that one of a,b,c,d is zero. □

Finally, in order for our discussion to be complete, remember from Cat 2.16 that we still have to discuss, both in 2D and 3D, the behavior of our theory under rotations, with most likely lots of interesting formulae here. However, the computations here can be quite complex, and we will defer the whole discussion to chapter 4 below, where we will kill everything with advanced tools, namely the Lorentz transform, and some geometry.

2e. Exercises

Exercises.

CHAPTER 3

Time and length

3a. Trains and clocks

We have seen so far that the Einstein principles, telling us that $c+v=c$, require summing the speeds in a tricky way, with the formula in 1D being as follows, in $c=1$ units, and with the general 3D formula appearing as a variation of this:

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

This is very nice, and in a better world, where time and distances are irrelevant, and only speed matters, that would be the end of what we have to say. Actually, as an interesting thought experiment, try to imagine what that better world would look like. Quite surprisingly, that is quite close to the modern world we live in:

(1) First, who cares about distances, what matters are time and speed. It is indeed far more pleasant to drive 30 minutes at 100 mph, than 1 hour at 50 mph.

(2) But then, thinking well, who cares about time too. It is indeed far more pleasant to drive 1 hour at 100 mph, than 45 minutes in a traffic jam.

This being said, forgetting about our modern world, and the fast and the furious, let us go back to our original goal, fixing classical mechanics. Speed is distance/time, so we must now fix as well distance, or time, or both. Fortunately the solution to the problem, involving a Gedankenexperiment using a train and a clock, is unique. Let us first discuss, following as usual Einstein, what happens to time. Here the result is as follows:

THEOREM 3.1. *Relativistic time is subject to Lorentz dilation*

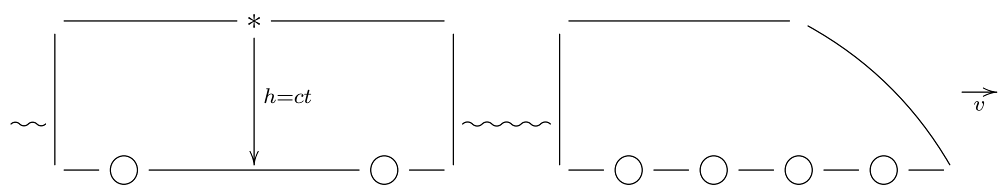
$$t \to \gamma t$$

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

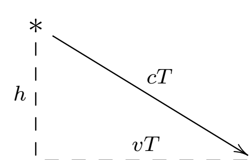
$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

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

PROOF. Assume indeed that we have a train, moving to the right with speed v, through vacuum. In order to compute the height h of the train, the passenger onboard switches on the ceiling light bulb, measures the time t that the light needs to hit the floor, by travelling at speed c, and concludes that the train height is $h = ct$:



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



Now by Pythagoras applied to this triangle, we have:

$$h^2 + (vT)^2 = (cT)^2$$

Thus, the observer on the ground will reach to the following formula for h:

$$h = \sqrt{c^2 - v^2} \cdot T$$

But h must be the same for both observers, so we have the following formula:

$$\sqrt{c^2 - v^2} \cdot T = ct$$

It follows that the two times t and T are indeed not equal, and are related by:

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

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

As a first comment, the above result is due to Einstein, as all the mathematics in this chapter. The occurrence of Lorentz, who previously discovered similar phenomena in the context of electromagnetism, is something very interesting, ultimately standing as a heavy and final piece of support for Einstein's theory, the idea being that "among Newton and Maxwell, the latter was right, and Einstein proved it". More on this later.

Let us discuss now what happens to length. Intuitively, since speed is distance/time, and since time gets dilated, we can somehow expect for distance to get dilated too.

However, this is wrong, and what happens, after due thinking and computations, is that distance gets in fact contracted, by the same factor, the result being as follows:

THEOREM 3.2. *Relativistic length is subject to Lorentz contraction*

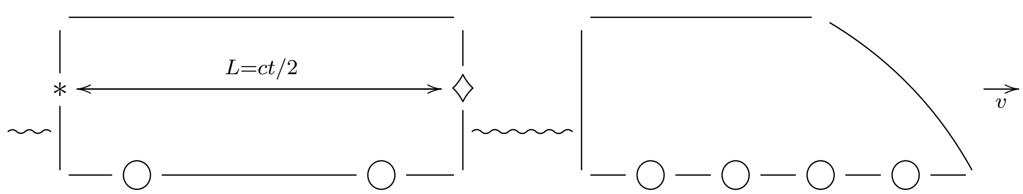
$$L \to L/\gamma$$

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

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

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

PROOF. As before in the proof of Theorem 3.1, meaning in the same train travelling at speed v, in vacuum, imagine now that the passenger wants to measure the length L of the car. For this purpose he switches on the light bulb, now at the rear of the car, and measures the time t needed for the light to reach the front of the car, and get reflected back by a mirror installed there, according to the following scheme:



He concludes that, as marked above, the length L of the car is given by:

$$L = \frac{ct}{2}$$

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

$$T = T_1 + T_2 = \frac{l}{c - v} + \frac{l}{c + v} = \frac{2lc}{c^2 - v^2}$$

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

$$\frac{2lc}{c^2 - v^2} = \gamma t = \frac{2\gamma L}{c}$$

Thus, the two lengths L and l are indeed not equal, and related by:

$$l = \frac{\gamma L(c^2 - v^2)}{c^2} = \gamma L\left(1 - \frac{v^2}{c^2}\right) = \frac{\gamma L}{\gamma^2} = \frac{L}{\gamma}$$

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

As before with time dilation, the above result is due to Einstein, with the occurrence of Lorentz coming from some previously discovered phenomena, of the same nature, in the context of electromagnetism. More on this later, in Part II of the present book, when talking unification of special relativity with electromagnetism.

3b. Bugs and fixes

What we have in Theorem 3.1 and Theorem 3.2 is certainly very interesting, and suggests doing a plethora of things, which can include, among others:

(1) Passing to 3D, and unifying what we have with the speed theory from chapter 2.

(2) Understanding the relation with the Lorentz findings from electromagnetism.

(3) Using time dilation as to escape cops, aging, death, and many other.

All this looks very interesting, and we will certainly explore all this, but before anything, are Theorem 3.1 and Theorem 3.2 really correct? I don't know about you, but personally I feel quite uneasy with the proof of Theorem 3.2, and I have on my mind:

QUESTION 3.3. *What is the meaning of $c+v$ from the proof of Theorem 3.2? And also, what is the meaning of $c-v$, from that same proof?*

And isn't this is a good question. Indeed, we perfectly know from the Einstein principles that talking about $c+v$, with the mathematical meaning $c+v>c$, as obviously employed in the proof of Theorem 3.2, is forbidden. As for the use of $c-v$, with the mathematical meaning $c-v<c$, as we did in that same proof, at the first glance this seems quite reasonable. But this is in fact not honest either, because we perfectly know, again from the Einstein principles, that the speed of light is c for all observers.

Summarizing, we are in trouble, and we must ask the cat. And cat says:

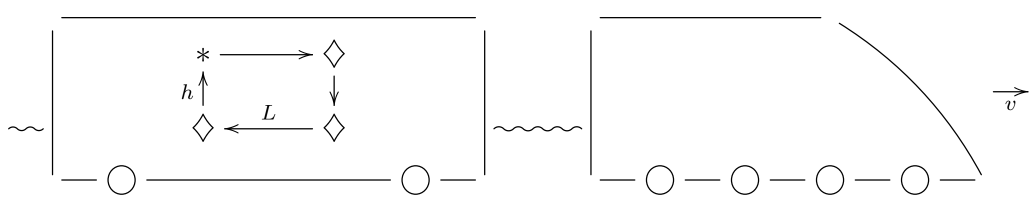
CAT 3.4. *That proof seems okay, at the advanced level $v>c$ is possible, and certainly so in an asymptotic sense. Change your diet, and you'll understand.*

Damn cat, this is not very useful. Not the good time for jokes. Although with cat you never know when he jokes or not. In fact, after observing him for a while, I'm pretty much sure that he knows something about $v>c$. But we're here for talking about relativity and $v<c$, right, so this kind of advanced cat stuff, hunting techniques for birds and mice, cosmological inflation, or whatever that is, will certainly not help.

So, leaving now cat alone, to play outside, at whatever speeds he likes, let us carefully examine Theorem 3.2 and its proof. Obviously, that is something quite clever, somewhat of Feynman flavor, and aren't we physicists in love with such things. But for being totally honest on this matter, we have now to fix that. But, how to do this?

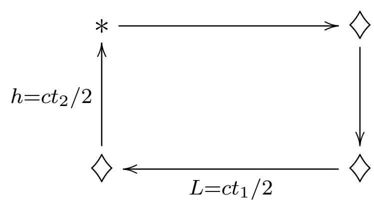
A first idea comes by comparing Theorem 3.1 and Theorem 3.2, and their proofs. Indeed, in both proofs there is telling about a train running at speed v, and then a light bulb being lit, first on the vertical, and then on the horizontal. But this suggests unifying Theorem 3.1 and Theorem 3.2, by using a rectangular setup, as follows:

QUESTION 3.5. *What happens when using inside our train running at speed v a rectangular setup, with 3 mirrors, as follows,*

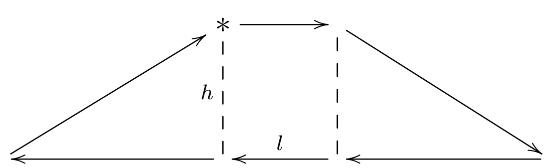


that is, when viewing this from the ground, in the non-degenerate case, $h, L > 0$, can we obtain both time dilation, $T = \gamma t$, and length contraction, $l = L/\gamma$?

Normally this looks quite reasonable, and from the perspective of the train passenger, the experiment will involve two lengths and two times, as follows:



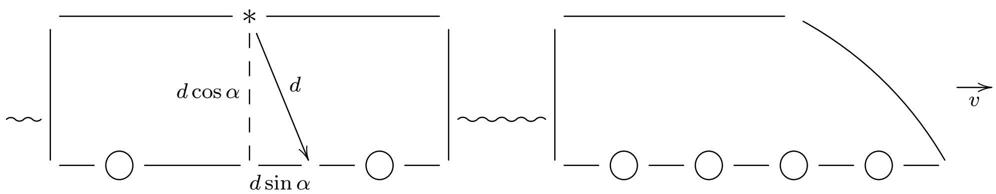
Regarding now the perspective from the ground, in view of what we already know about the limiting cases $h, L = 0$, this should normally look as follows:



However, this remains something quite formal, and not really innovating, and when analyzing this, we have all chances to get back to exactly the same difficulties that we met in the proof of Theorem 3.2. So, as a conclusion, this is most likely a bad idea.

Along the same lines, namely unifying Theorem 3.1 and Theorem 3.2, we can potentially do this as well by using a triangular setup, as follows:

QUESTION 3.6. *What happens when using a bulb emitting at an angle $\alpha \in \mathbb{R}$,*



viewed from the ground?

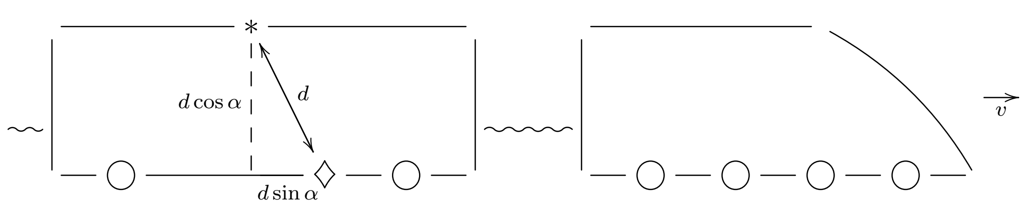
And this looks quite interesting. Indeed, we know how to solve this question for $\alpha = 0$, as explained in the proof of Theorem 3.1, with the end result being Theorem 3.1 itself, Lorentz dilation of time. We also sort of know how to solve this question for the limiting case $\alpha = \pi/2$, as explained in the proof of Theorem 3.2, with the end result being Theorem 3.2 itself, Lorentz contraction of length. So, our strategy will be to generalize the proof of Theorem 3.1, as to deal with arbitrary angles $\alpha \in [0, \pi/2)$, in a clean way, and then, in order to clarify what happens with Theorem 3.2, pass our results to the limit:

$$\alpha \to \frac{\pi}{2}$$

And this sounds good, doesn't it. Expect some 2D calculus, and then some asymptotics to be worked out. By the way, speaking asymptotics, wasn't about asymptotics that cat was talking, in his cryptic advice 3.4. So good, we're getting wiser every day.

Getting started now, the first question is whether we want to use a mirror, or not. Since the proof of Theorem 3.1 looks easy to upgrade with a mirror, while the proof of Theorem 3.2 does not make much sense without a mirror, we will conclude that the mirror is an essential device, and so, we will use one. So, let us upgrade Question 3.6 into:

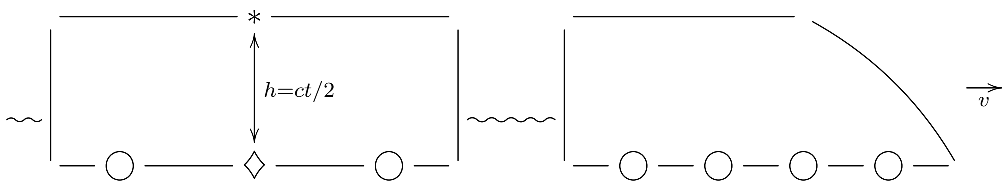
QUESTION 3.7. *What happens when using a bulb emitting at an angle $\alpha \in [0, \pi/2)$,*



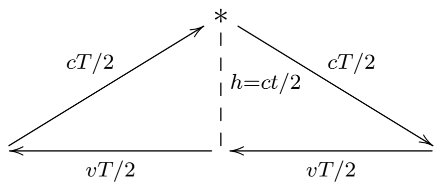
and reflected by a mirror, as above, when viewed from the ground?

In order now to solve this question, let us first look at the simplest case, $\alpha = 0$. Here we already know the answer, from Theorem 3.1, and by upgrading the proof there, by using a mirror, as above, we are led to the following conclusion:

ANSWER 3.8. *In the case $\alpha = 0$, the view from the train, namely*

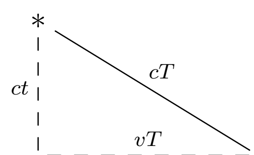


translates into the following view from the ground,



and leads to Lorentz time dilation, $T = \gamma t$, via Pythagoras.

To be more precise, regarding the last assertion, the triangle on the right is similar, by a factor of 1/2, to the triangle that we used in the proof of Theorem 3.1, namely:



Thus, we are led via Pythagoras to Lorentz time dilation, $T = \gamma t$.

As a second piece of data, although a bit conjectural, let us record as well what happens in the limiting case, $\alpha = \pi/2$. Here we know from Theorem 3.2 and its proof, regardless of what the precise status of that proof is, that viewed from the ground, we have to deal twice with the length of the car l observed from the ground, as follows:

$$l = (c - v)T_1 = (c + v)T_2$$

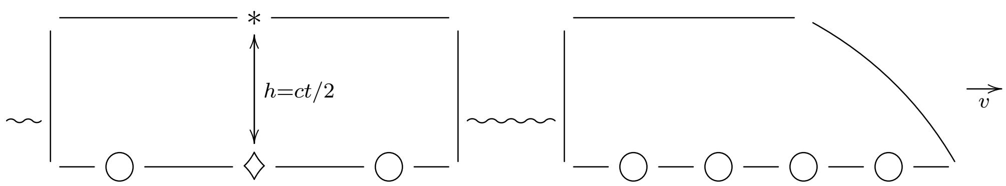
To be more precise, passed the uncertainties with the meaning of $c \pm v$, here the length is $l = L/\gamma$, and T_1, T_2 are the formal forward and backwards times, given by:

$$T_1 = \frac{l}{c - v} = \frac{L}{\gamma(c - v)} \quad , \quad T_2 = \frac{l}{c + v} = \frac{L}{\gamma(c + v)}$$

It is however a bit unclear how to include these times T_1, T_2 on a triangular diagram as in Answer 3.8, with the problem coming from the fact that at $\alpha = \pi/2$ this triangular diagram can only be degenerate, with the 2 lengths being proportional to T_1, T_2.

Summarizing, we are a bit stuck, with the problem coming from the fact that we would like to avoid triangular diagrams as in Answer 3.8, and the Pythagoras theorem. But, to any problem its solution. The problem comes from Answer 3.8, so let us formulate:

QUESTION 3.9. *In the case $\alpha = 0$, where the view from the train is*



can we recover Lorentz dilation, $T = \gamma t$, without triangles and Pythagoras?

And the point now is that we have in our bag an alternative tool indeed, for dealing with such questions, namely the Einstein speed summation formulae from chapter 2. So, let us get now into this, answering Question 3.9 by using that formulae.

Let us first compute the downwards speed of light, observed from the ground. We denote by Greek letters the 3D speed vectors involved, in $c = 1$ units, as follows:

$$\nu = \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix} \quad , \quad \omega = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}$$

We have then $< \nu, \omega >= 0$, and the following vector product formulae:

$$\nu \times \omega = \begin{pmatrix} 0 \\ 0 \\ -v \end{pmatrix} \quad , \quad \nu \times (\nu \times \omega) = \begin{pmatrix} 0 \\ v^2 \\ 0 \end{pmatrix}$$

In order to formulate now the results, it is convenient to get back to 2D, by cutting the last 0 components of all vectors involved. That is, we make the following reset:

$$\nu = \begin{pmatrix} v \\ 0 \end{pmatrix} \quad , \quad \omega = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

With this convention, the composed speed vector is as follows, in $c = 1$ units:

$$\nu +_e \omega = \begin{pmatrix} v \\ -1 \end{pmatrix} + \frac{1}{1 + \sqrt{1 - v^2}} \begin{pmatrix} 0 \\ v^2 \end{pmatrix}$$

However, things are not over here, and we can do better. Indeed, observe that we have the following formula, coming from definitions:

$$\frac{v^2}{1 + \sqrt{1 - v^2}} = 1 - \sqrt{1 - v^2}$$

Thus, the composed speed vector computed above simply becomes:

$$\nu +_e \omega = \begin{pmatrix} v \\ -\sqrt{1 - v^2} \end{pmatrix}$$

We can now recover Theorem 3.1, in a better way. Indeed, the observed light speed from the ground is $\nu +_e \omega$, having as vertical component the following quantity:

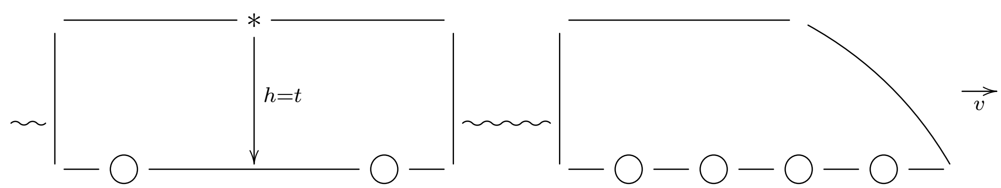
$$(\nu +_e \omega)_2 = -\sqrt{1-v^2}$$

Observe that this is smaller in absolute value than $\omega_2 = -1$. Now since the vertical distance travelled is the same viewed from the train and from the ground, we conclude that, when viewing things from the ground, time dilates by the following factor:

$$\gamma = \frac{\omega_2}{(\nu +_e \omega)_2} = \frac{-1}{-\sqrt{1-v^2}} = \frac{1}{\sqrt{1-v^2}}$$

But this is exactly the Lorentz factor, written in $c = 1$ units. So, let us record:

Answer 3.10. *In the vertical case, where $\alpha = 0$, and without mirror,*



the observed light speed from the ground is given by the formula

$$\nu +_e \omega = \begin{pmatrix} v \\ -\sqrt{1-v^2} \end{pmatrix}$$

and looking at the vertical component, we recover Lorentz time dilation, namely:

$$T = \gamma t$$

Moreover, the same computation can be performed upwards, or with a mirror.

Here the first assertion is something that we know from the above. As for the last assertion, when doing exactly the same experiment with the light travelling upwards, that is, with the bulb being now lit on the floor, the formulae established above remain valid, by modifying the -1 component of ω into a 1 component, and we get:

$$\nu +_e (-\omega) = \begin{pmatrix} v \\ \sqrt{1-v^2} \end{pmatrix}$$

Thus, as before, we conclude that time gets dilated by the following factor:

$$\gamma = \frac{(-\omega)_2}{(\nu +_e (-\omega))_2} = \frac{1}{\sqrt{1-v^2}}$$

Finally, we can merge if we want these two computations, by doing the experiment with a mirror, and we obtain the same thing in the end, namely:

$$T = \gamma t$$

Before moving forward, as an instructive computation, let us see as well what happens when attempting to use instead $\omega +_e \nu$ and $(-\omega) +_e \nu$. Our vectors are:

$$\omega = \begin{pmatrix}0\\-1\\0\end{pmatrix} \quad , \quad \nu = \begin{pmatrix}v\\0\\0\end{pmatrix}$$

We have then $< \omega, \nu >= 0$, and the following vector product formulae:

$$\omega \times \nu = \begin{pmatrix}0\\0\\v\end{pmatrix} \quad , \quad \omega \times (\omega \times \nu) = \begin{pmatrix}-v\\0\\0\end{pmatrix}$$

By cutting the last component, as before, the composed speed vector is as follows:

$$\omega +_e \nu = \begin{pmatrix}v\\-1\end{pmatrix} + \begin{pmatrix}-v\\0\end{pmatrix} = \begin{pmatrix}0\\-1\end{pmatrix}$$

Similarly, the upwards observed speed is given by the following formula:

$$(-\omega) +_e \nu = \begin{pmatrix}v\\1\end{pmatrix} + \begin{pmatrix}-v\\0\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}$$

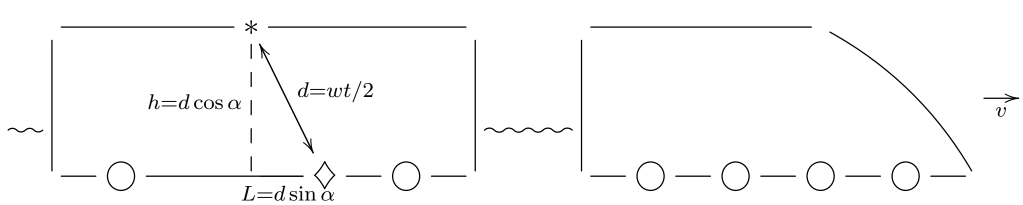
These formulae are not surprising, in view of what we know from chapter 2, and of course worth nothing in relation with our problem. But, let us formulate:

CONCLUSION 3.11. *Never mess up the composed speeds*

$$\nu +_e \omega \quad , \quad \omega +_e \nu$$

in 2 *dimensions or more.*

With this done, let us get now to Question 3.7 in general. We recall that the setup there was as follows, with a light bulb emitting at an angle $\alpha \in [0, \pi/2)$:



In order to see what happens, viewed from the ground, we must do the same computations as above, downwards and upwards, with an angle α added. For the downwards computation, the 3D speed vectors are as follows, with the same conventions as before:

$$\nu = \begin{pmatrix}v\\0\\0\end{pmatrix} \quad , \quad \omega = \begin{pmatrix}\sin\alpha\\-\cos\alpha\\0\end{pmatrix}$$

The scalar product of these vectors is given by the following formula:

$$< \nu, \omega >= v \sin \alpha$$

Also, we succesively have the following vector product formulae:

$$\nu \times \omega = \begin{pmatrix} 0 \\ 0 \\ -v \cos \alpha \end{pmatrix} \quad , \quad \nu \times (\nu \times \omega) = \begin{pmatrix} 0 \\ v^2 \cos \alpha \\ 0 \end{pmatrix}$$

In order to formulate now the result, it is convenient to get back to 2D, by cutting the last 0 components of all vectors involved. That is, we make the following reset:

$$\nu = \begin{pmatrix} v \\ 0 \end{pmatrix} \quad , \quad \omega = \begin{pmatrix} \sin \alpha \\ -\cos \alpha \end{pmatrix}$$

With this convention, the composed speed vector is as follows, in $c = 1$ units:

$$\begin{aligned} \nu +_e \omega &= \frac{1}{1 + v \sin \alpha} \left[\begin{pmatrix} v + \sin \alpha \\ -\cos \alpha \end{pmatrix} + \frac{1}{1 + \sqrt{1 - v^2}} \begin{pmatrix} 0 \\ v^2 \cos \alpha \end{pmatrix} \right] \\ &= \frac{1}{1 + v \sin \alpha} \left[\begin{pmatrix} v + \sin \alpha \\ -\cos \alpha \end{pmatrix} + (1 - \sqrt{1 - v^2}) \begin{pmatrix} 0 \\ \cos \alpha \end{pmatrix} \right] \\ &= \frac{1}{1 + v \sin \alpha} \begin{pmatrix} v + \sin \alpha \\ -\sqrt{1 - v^2} \cdot \cos \alpha \end{pmatrix} \end{aligned}$$

Similarly, the upwards observed speed is given by the following formula:

$$\nu +_e (-\omega) = \frac{1}{1 - v \sin \alpha} \begin{pmatrix} v - \sin \alpha \\ \sqrt{1 - v^2} \cdot \cos \alpha \end{pmatrix}$$

With these formulae in hand, let us first discuss what happens on the vertical. Here the observed speeds, downwards and upwards, are as follows:

$$S_1 = -\frac{\sqrt{1 - v^2} \cdot \cos \alpha}{1 + v \sin \alpha} \quad , \quad S_2 = \frac{\sqrt{1 - v^2} \cdot \cos \alpha}{1 - v \sin \alpha}$$

Since the vertical distance travelled is the same, namely h, we get:

$$T_1 = \frac{h}{|S_1|} = h \cdot \frac{1 + v \sin \alpha}{\sqrt{1 - v^2} \cdot \cos \alpha} = \gamma h \cdot \frac{1 + v \sin \alpha}{\cos \alpha}$$

$$T_2 = \frac{h}{S_2} = h \cdot \frac{1 - v \sin \alpha}{\sqrt{1 - v^2} \cdot \cos \alpha} = \gamma h \cdot \frac{1 - v \sin \alpha}{\cos \alpha}$$

Now since $h = d \cos \alpha$, with $d = t/2$, we obtain the following formulae:

$$T_1 = \gamma \cdot \frac{t \cos \alpha}{2} \cdot \frac{1 + v \sin \alpha}{\cos \alpha} = (1 + v \sin \alpha) \frac{\gamma t}{2}$$

$$T_2 = \gamma \cdot \frac{t \cos \alpha}{2} \cdot \frac{1 - v \sin \alpha}{\cos \alpha} = (1 - v \sin \alpha) \frac{\gamma t}{2}$$

But these formulae are quite nice, containing exactly what we need, and we can therefore formulate, as a partial answer to Question 3.7:

ANSWER 3.12. *When looking on the vertical, in the context of Question 3.7, we conclude that the total time $T = T_1 + T_2$ is subject to Lorentz dilation, namely:*

$$T = \gamma t$$

However, the partial times T_1, T_2 are not related to $t/2$ by a simple formula.

As an illustration for the above formulae of T_1, T_2, which are quite tricky, in the case $\alpha = 0$ we obtain $T_1 = T_2 = \gamma t/2$, which is correct. Also, in the case $\alpha = \pi/2$ we obtain the partial times previously computed in the proof of Theorem 3.2, namely:

$$T_1 = (1+v)\frac{\gamma t}{2} = (1+v)\gamma L = \frac{L}{\gamma(1-v)}$$

$$T_2 = (1-v)\frac{\gamma t}{2} = (1-v)\gamma L = \frac{L}{\gamma(1+v)}$$

Now let us discuss what happens on the horizontal. At $\alpha = \pi/2$, as in the proof of Theorem 3.2, we formally have the following formulae, each giving $l = L/\gamma$:

$$T_1 = \frac{l}{1-v} \quad , \quad T_2 = \frac{l}{1+v}$$

However, as already mentioned on several occasions, these formulae are something truly formal, not making much sense, when thinking a bit at their meaning.

So, our purpose in what follows will be that of finding another proof of the Lorentz contraction formula, $l = L/\gamma$. For this purpose, let us go back to the observed speeds computed before. The observed speeds, forward and backwards, are as follows:

$$V_1 = \frac{v + \sin\alpha}{1 + v\sin\alpha} \quad , \quad V_2 = \frac{v - \sin\alpha}{1 - v\sin\alpha}$$

As an illustration here, at $\alpha = 0$ we obtain $V_1 = V_2 = v$, which is correct. Also, at $\alpha = \pi/2$ we obtain the following formulae, which are correct too:

$$V_1 = \frac{v+1}{1+v} = 1 \quad , \quad V_2 = \frac{v-1}{1-v} = -1$$

The problem now is that of understanding what happens on the horizontal. For this purpose, our first claim is that we have the following key formula:

$$V_1 T_1 + V_2 T_2 = vT$$

Normally this comes from the fact that both the above expressions compute the horizontal displacement of the bulb, in the observed time from the ground T. But we can check this too by using our formulae for all the quantities involved. Indeed, we have:

$$V_1T_1 = \frac{v+\sin\alpha}{1+v\sin\alpha}\cdot(1+v\sin\alpha)\frac{\gamma t}{2} = (v+\sin\alpha)\frac{\gamma t}{2}$$

$$V_2T_2 = \frac{v-\sin\alpha}{1-v\sin\alpha}\cdot(1-v\sin\alpha)\frac{\gamma t}{2} = (v-\sin\alpha)\frac{\gamma t}{2}$$

Thus by summing we obtain the formula claimed above, namely:

$$V_1T_1 + V_2T_2 = v\gamma T = vT$$

Our goal now will be that of fine-tuning this formula, as to make appear the lengths l, L, hopefully related by the Lorentz contraction formula $l = L/\gamma$. In order to do so, observe that by using $L = d\sin\alpha$ with $d = t/2$, we have:

$$V_1T_1 = \frac{v\gamma t}{2} + \frac{\sin\alpha\cdot\gamma t}{2} = \frac{vT}{2} + \gamma L$$

$$V_2T_2 = \frac{v\gamma t}{2} - \frac{\sin\alpha\cdot\gamma t}{2} = \frac{vT}{2} - \gamma L$$

Thus, we have here a better explanation of our formula $V_1T_1 + V_2T_2 = vT$, coming from the following beautiful looking formulae:

$$V_1T_1 = \frac{vT}{2} + \gamma L \quad , \quad V_2T_2 = \frac{vT}{2} - \gamma L$$

The problem however is that, curiously, the Lorentz factors γ appear with the wrong exponent, I mean life would have been so much simpler if we had on the right L/γ instead of γL, we could have probably turned this whole thing quickly into $l = L/\gamma$.

Struggling with the above formulae, in order to find their correct interpretation in terms of lengths, or why not finding the mistakes inside them, but they are in fact both correct, can easily take hours or even days, so time to ask the cat. And cat says:

Cat 3.13. *Don't mix partial times, unless you are at the advanced level.*

Thanks cat, so if I understand well the idea of quickly removing $vT/2$ from the quantities V_1T_1, V_2T_2, although quite natural, is something not very good. So, let us try instead to remove vT_1, vT_2 from these quantities, see what we get. We first have:

$$\begin{aligned}
V_1T_1 - vT_1 &= (v+\sin\alpha)\frac{\gamma t}{2} - v(1+v\sin\alpha)\frac{\gamma t}{2}\\
&= \frac{\gamma t}{2}(v+\sin\alpha - v - v^2\sin\alpha)\\
&= \frac{\gamma t}{2}(\sin\alpha - v^2\sin\alpha)\\
&= \frac{\gamma t}{2}\cdot\sin\alpha\cdot(1-v^2)\\
&= \frac{\gamma t}{2}\cdot\sin\alpha\cdot\frac{1}{\gamma^2}\\
&= \frac{1}{\gamma}\cdot\frac{t\sin\alpha}{2}\\
&= \frac{1}{\gamma}\cdot d\sin\alpha\\
&= \frac{L}{\gamma}
\end{aligned}$$

Similarly, or just by using $V_1T_1 + V_2T_2 = vT$, we have as well:

$$V_2T_2 - vT_2 = -\frac{L}{\gamma}$$

Thus, we have now an alternative explanation for $V_1T_1 + V_2T_2 = vT$, coming from the following formulae, which are even more beautiful than those that we had before:

$$V_1T_1 = vT_1 + \frac{L}{\gamma} \quad , \quad V_2T_2 = vT_2 - \frac{L}{\gamma}$$

But this is exactly what we need, because thinking at what happens, when viewed from the ground, the quantities L/γ on the right must be the observed horizontal length from the ground l. Thus, we can now complement Answer 3.12 with:

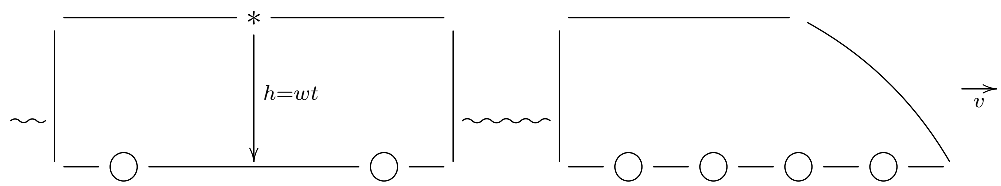
Answer 3.14. *When looking on the horizontal, in the context of Question 3.7, we conclude that the length is subject to Lorentz contraction, namely $l = L/\gamma$.*

As a conclusion, understanding the precise, rigorous functioning of the Lorentz dilation and contraction via trains and clocks is in fact a quite complicated business, needing days of computations in a row, and help from a cat, on the bottom line.

3c. Slow bulbs

We have seen that many things can be said, simply by using a Gedankenexperiment involving a train, a bulb, a mirror and a clock. So, let us further explore what we can do, with these simple tools. As a first interesting question, we have:

QUESTION 3.15. *What happens when using a slow bulb, emitting at $w < c$,*



viewed from the ground?

In order to discuss this, let us compute speed sums first, by using the formulae from chapter 2. We denote by Greek letters the 3D speed vectors, as follows:

$$\nu = \begin{pmatrix} v\\ 0\\ 0\end{pmatrix} \quad , \quad \omega = \begin{pmatrix} 0\\ -w\\ 0\end{pmatrix}$$

We have then $<\nu,\omega>=0$, and the following vector product formulae:

$$\nu\times\omega = \begin{pmatrix} 0\\ 0\\ -vw\end{pmatrix} \quad , \quad \nu\times(\nu\times\omega) = \begin{pmatrix} 0\\ v^2w\\ 0\end{pmatrix}$$

In order to formulate now the results, it is convenient to get back to 2D, by cutting the last 0 components of all vectors involved. That is, we make the following reset:

$$\nu = \begin{pmatrix} v\\ 0\end{pmatrix} \quad , \quad \omega = \begin{pmatrix} 0\\ -w\end{pmatrix}$$

With this convention, the composed speed vector is as follows, in $c = 1$ units:

$$\begin{aligned}\nu +_e \omega &= \begin{pmatrix} v\\ -w\end{pmatrix} + \frac{1}{1+\sqrt{1-v^2}}\begin{pmatrix} 0\\ v^2w\end{pmatrix}\\ &= \begin{pmatrix} v\\ -w\end{pmatrix} + (1-\sqrt{1-v^2})\begin{pmatrix} 0\\ w\end{pmatrix}\\ &= \begin{pmatrix} v\\ -\sqrt{1-v^2}\,w\end{pmatrix}\end{aligned}$$

We can now recover Theorem 3.1, in a more solid way. Indeed, the observed light speed from the ground is $\nu +_e \omega$, having as vertical component the following quantity:

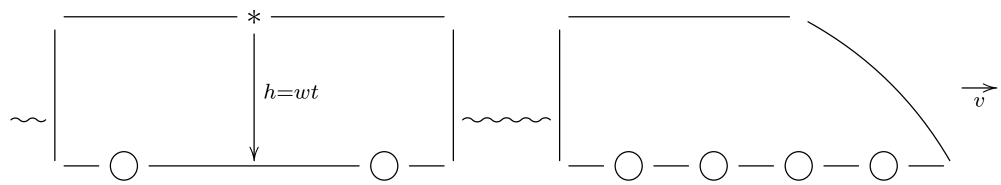
$$(\nu +_e \omega)_2 = -\sqrt{1-v^2}\,w$$

Observe that this is smaller in absolute value than $\omega_2 = -w$. Now since the vertical distance travelled is the same viewed from the train and from the ground, we conclude that, when viewing things from the ground, time dilates by the following factor:

$$\gamma = \frac{\omega_2}{(\nu +_e \omega)_2} = \frac{-w}{-\sqrt{1-v^2}\,w} = \frac{1}{\sqrt{1-v^2}}$$

But this is exactly the Lorentz factor, written in $c = 1$ units. So, let us record:

ANSWER 3.16. *What using a slow bulb, emitting vertically at $w < c$,*

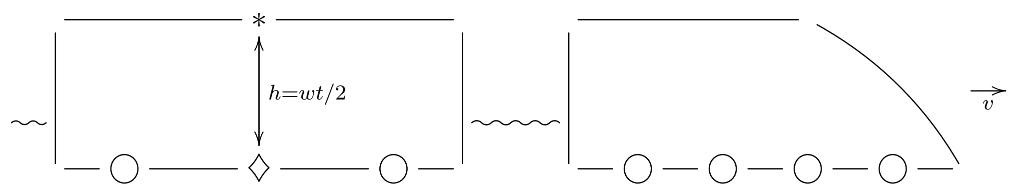


the observed light speed from the ground is given by the formula

$$\nu +_e \omega = \begin{pmatrix} v \\ -\sqrt{1-v^2}\, w \end{pmatrix}$$

and looking at the vertical component, we recover Lorentz time dilation, $T = \gamma t$.

Before going ahead, with the slow bulb emitting at an angle α, let us record what happens to the above computation when adding a mirror, as follows:



Here the observed light speeds from the ground, downwards and upwards, are:

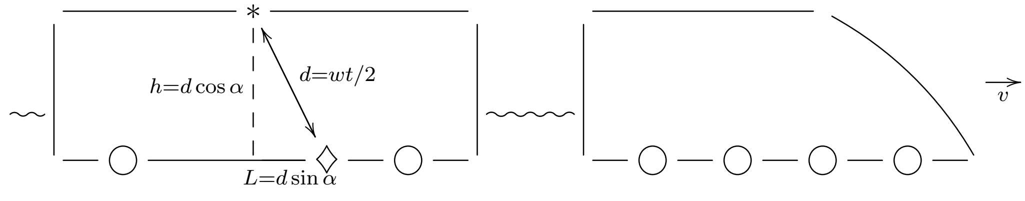
$$\nu +_e \omega = \begin{pmatrix} v \\ -\sqrt{1-v^2}\, w \end{pmatrix}$$

$$\nu +_e (-\omega) = \begin{pmatrix} v \\ \sqrt{1-v^2}\, w \end{pmatrix}$$

Thus, by looking at the vertical component, either on the downward path, or on the upwards one, or on the combination of these, we recover Lorentz time dilation, $T = \gamma t$.

More generally now, we have the following generalization of Question 3.15:

QUESTION 3.17. *What happens when using a slow bulb emitting at speed $w < c$, and at an angle $\alpha \in [0, \pi/2)$, reflected by a mirror,*



when viewed from the ground?

As before with other such questions, at the first glance this might look like something quite technical. But after thinking a bit, solving this question appears to be of crucial importance, because we are more or less dealing here, and with more intuitive tools than in chapter 2, with the general summation problem for speeds, in general relativity.

In order to solve now our question, we must do the same computation as for Question 3.15, downwards and upwards, with an angle α added. For the downwards computation, the 3D speed vectors are as follows, with the same conventions as before:

$$\nu=\begin{pmatrix}v\\0\\0\end{pmatrix}\quad,\quad\omega=\begin{pmatrix}w\sin\alpha\\-w\cos\alpha\\0\end{pmatrix}$$

We have then $<\nu,\omega>=vw\sin\alpha$, and the following vector product formulae:

$$\nu\times\omega=\begin{pmatrix}0\\0\\-vw\cos\alpha\end{pmatrix}\quad,\quad\nu\times(\nu\times\omega)=\begin{pmatrix}0\\v^2w\cos\alpha\\0\end{pmatrix}$$

In order to formulate now the result, it is convenient to get back to 2D, by cutting the last 0 components of all vectors involved. That is, we make the following reset:

$$\nu=\begin{pmatrix}v\\0\end{pmatrix}\quad,\quad\omega=\begin{pmatrix}w\sin\alpha\\-w\cos\alpha\end{pmatrix}$$

With this convention, the composed speed vector is as follows, in $c=1$ units:

$$\begin{aligned}\nu+_e\omega&=\frac{1}{1+vw\sin\alpha}\left[\begin{pmatrix}v+w\sin\alpha\\-w\cos\alpha\end{pmatrix}+\frac{1}{1+\sqrt{1-v^2}}\begin{pmatrix}0\\v^2w\cos\alpha\end{pmatrix}\right]\\&=\frac{1}{1+vw\sin\alpha}\left[\begin{pmatrix}v+w\sin\alpha\\-w\cos\alpha\end{pmatrix}+(1-\sqrt{1-v^2})\begin{pmatrix}0\\w\cos\alpha\end{pmatrix}\right]\\&=\frac{1}{1+vw\sin\alpha}\begin{pmatrix}v+w\sin\alpha\\-\sqrt{1-v^2}\cdot w\cos\alpha\end{pmatrix}\end{aligned}$$

Similarly, or just by replacing $w\to-w$, the upwards observed speed is:

$$\nu+_e(-\omega)=\frac{1}{1-vw\sin\alpha}\begin{pmatrix}v-w\sin\alpha\\\sqrt{1-v^2}\cdot w\cos\alpha\end{pmatrix}$$

With these formulae in hand, let us first discuss what happens on the vertical. Here the observed speeds, downwards and upwards, are as follows:

$$S_1=-\frac{\sqrt{1-v^2}\cdot w\cos\alpha}{1+vw\sin\alpha}$$

$$S_2=\frac{\sqrt{1-v^2}\cdot w\cos\alpha}{1-vw\sin\alpha}$$

Since the vertical distance travelled is the same, namely h, we get:

$$T_1 = \frac{h}{|S_1|} = h \cdot \frac{1 + vw\sin\alpha}{\sqrt{1-v^2}\cdot w\cos\alpha} = \gamma h \cdot \frac{1 + vw\sin\alpha}{w\cos\alpha}$$

$$T_2 = \frac{h}{S_2} = h \cdot \frac{1 - vw\sin\alpha}{\sqrt{1-v^2}\cdot w\cos\alpha} = \gamma h \cdot \frac{1 - vw\sin\alpha}{w\cos\alpha}$$

Now since $h = d\cos\alpha$, with $d = wt/2$, we obtain the following formulae:

$$T_1 = \gamma \cdot \frac{wt\cos\alpha}{2} \cdot \frac{1 + vw\sin\alpha}{w\cos\alpha} = (1 + vw\sin\alpha)\frac{\gamma t}{2}$$

$$T_2 = \gamma \cdot \frac{wt\cos\alpha}{2} \cdot \frac{1 - vw\sin\alpha}{w\cos\alpha} = (1 - vw\sin\alpha)\frac{\gamma t}{2}$$

But these formulae are quite nice, containing exactly what we need, and we can therefore formulate, as a partial answer to Question 3.17:

Answer 3.18. *When looking on the vertical, in the context of Question 3.17, we conclude that the total time $T = T_1 + T_2$ is subject to Lorentz dilation, namely:*

$$T = \gamma t$$

However, the partial times T_1, T_2 are not related to $t/2$ by a simple formula.

As an illustration for the above formulae of T_1, T_2, which are quite tricky, in the case $\alpha = 0$ we obtain $T_1 = T_2 = \gamma t/2$, which is correct. Also, in the case $\alpha = \pi/2, w = 1$ we obtain the partial times previously computed in the proof of Theorem 3.2, namely:

$$T_1 = (1+v)\frac{\gamma t}{2} = (1+v)\gamma L = \frac{L}{\gamma(1-v)}$$

$$T_2 = (1-v)\frac{\gamma t}{2} = (1-v)\gamma L = \frac{L}{\gamma(1+v)}$$

Now let us discuss what happens on the horizontal. For this purpose, let us go back to the speeds computed before. The observed speeds, forward and backwards, are:

$$V_1 = \frac{v + w\sin\alpha}{1 + vw\sin\alpha} \quad , \quad V_2 = \frac{v - w\sin\alpha}{1 - vw\sin\alpha}$$

As an illustration here, at $\alpha = 0$ we obtain $V_1 = V_2 = v$, which is correct. Also, at $\alpha = \pi/2$ we obtain the following formulae, which are correct too:

$$V_1 = \frac{v + w}{1 + vw} \quad , \quad V_2 = \frac{v - w}{1 - vw}$$

The problem now is that of understanding what happens on the horizontal. For this purpose, our first claim is that we have the following key formula:

$$V_1T_1 + V_2T_2 = vT$$

Normally this comes from the fact that both the above expressions compute the horizontal displacement of the bulb, in the observed time from the ground T. But we can check this too by using our formulae for all the quantities involved. Indeed, we have:

$$V_1T_1 = \frac{v + w\sin\alpha}{1 + vw\sin\alpha} \cdot (1 + vw\sin\alpha)\frac{\gamma t}{2} = (v + w\sin\alpha)\frac{\gamma t}{2}$$

$$V_2T_2 = \frac{v - w\sin\alpha}{1 - vw\sin\alpha} \cdot (1 - vw\sin\alpha)\frac{\gamma t}{2} = (v - w\sin\alpha)\frac{\gamma t}{2}$$

Thus by summing we obtain the formula claimed above, namely:

$$V_1T_1 + V_2T_2 = v\gamma T = vT$$

Our goal now will be that of fine-tuning this formula, as to make appear the lengths l, L, hopefully related by the Lorentz contraction formula $l = L/\gamma$. In order to do so, observe that by using $L = d\sin\alpha$ with $d = wt/2$, we have:

$$V_1T_1 = \frac{v\gamma t}{2} + \frac{w\sin\alpha\gamma t}{2} = \frac{vT}{2} + \gamma L$$

$$V_2T_2 = \frac{v\gamma t}{2} - \frac{w\sin\alpha\gamma t}{2} = \frac{vT}{2} - \gamma L$$

Thus, we have here a better explanation of our formula $V_1T_1 + V_2T_2 = vT$, coming from the following beautiful looking formulae:

$$V_1T_1 = \frac{vT}{2} + \gamma L \quad , \quad V_2T_2 = \frac{vT}{2} - \gamma L$$

However, as before in the case $w = 1$, these formulae are a trap, and we can do better. Indeed, as explained in Cat 3.13, mixing partial times is a bad idea. So, let us try instead to remove vT_1, vT_2 from these quantities, see what we get. We first have:

$$\begin{aligned}
V_1T_1 - vT_1 &= (v + w\sin\alpha)\frac{\gamma t}{2} - v(1 + vw\sin\alpha)\frac{\gamma t}{2} \\
&= \frac{\gamma t}{2}(v + w\sin\alpha - v - v^2 w\sin\alpha) \\
&= \frac{\gamma t}{2}(w\sin\alpha - v^2 w\sin\alpha) \\
&= \frac{\gamma t}{2}\cdot w\sin\alpha(1 - v^2) \\
&= \frac{\gamma t}{2}\cdot w\sin\alpha\cdot\frac{1}{\gamma^2} \\
&= \frac{1}{\gamma}\cdot\frac{wt\sin\alpha}{2} \\
&= \frac{1}{\gamma}\cdot d\sin\alpha \\
&= \frac{L}{\gamma}
\end{aligned}$$

Similarly, or just by using $V_1T_1 + V_2T_2 = vT$, we have as well:

$$V_2T_2 - vT_2 = -\frac{L}{\gamma}$$

Thus, we have now an alternative explanation for $V_1T_1 + V_2T_2 = vT$, coming from:

$$V_1T_1 = vT_1 + \frac{L}{\gamma} \quad , \quad V_2T_2 = vT_2 - \frac{L}{\gamma}$$

But this is exactly what we need, because the quantities L/γ on the right must be the observed length from the ground l. Thus, we can now complement Answer 3.18 with:

Answer 3.19. *When looking on the horizontal, in the context of Question 3.17, we conclude that the length is subject to Lorentz contraction, namely $l = L/\gamma$.*

Summarizing, we made it, but all this shows that the mastering of relativity theory via trains and clocks is a rather complicated business. We will be back to this in the next chapter, with a more clever way of approaching such questions.

3d. Relativistic tricks

Relativistic tricks, for escaping cops, aging, and many other.

3e. Exercises

Exercises.

CHAPTER 4

Frame change

4a. Lorentz transformation

Welcome to relativity theory, take two. We kept the best of this introductory Part I for the end. After the 50 pages of computations that we just did, and all the sweat, blood and tears coming with that, we can agree, I hope, on the following principle:

PRINCIPLE 4.1 (Einstein). *Space and time are related.*

This is of course something advanced, utterly contradicting our intuition, and everything that we learned since childhood. However, we built so much evidence for this, with the experiments of Fizeau and Michelson-Morley, then following Einstein, with all sorts of bizarre speed summation formulae, and then again following Einstein with the Lorentz dilation of time, and contraction of length, that we will have to adopt this.

Importantly, this principle can only contradict our religious beliefs too, because even when accepting that for us, poor humans, space and time are related, what about God. Isn't he supposed to be free as a bird, knowing in real time everything that happens in the Universe, and not bound by $v < c$ or any other space and time constraints.

This is of course rather personal matter, and personally I think that Principle 4.1 only applies to the physics which is visible to us, humans. We are kind of slow, aren't we, and my cat, or yours, can only confirm this, so we can see only "slow physics", and so finally, Principle 4.1 is something normal. By the way, as an interesting thing here, the Big Bang as we know it cannot be explained via slow physics, so working on that concepts, namely cosmic inflation, dark energy and matter, and so on, can even give us a taste of what "fact physics" might mean. Although, again personally, I'd rather label that as "super-fast physics", with fast being reserved for what my cat knows.

Anyway, too much talking, so getting to the point now, Principle 4.1, once deeply agreed upon, is all that we need for developing relativity theory, meaning quickly rewriting all that we know, and then doing many more. Let us formulate a goal, as follows:

GOAL 4.2. *Find the formula of space-time frame change in relativity, then trivialize all that we know with this, and then do many more.*

As a first remark, this might look easy, but expect however some abstract mathematics to be involved. Indeed, once we will get into such things, gone our usual 3D frames for viewing everything, we will have to see and work in 4D, with (x, y, z, t) being our coordinates. And with this being of course among the many reasons for which we deferred this discussion to this final chapter of this introductory Part I of the present book.

In practice now, in order to reach to a precise formula for the space-time frame change, we must first discuss the 3D extension of the material that we learned in chapter 3, concerning the Lorentz dilation of time, and the Lorentz contraction of length. And, good news here, there is basically just some straightforward math to be done, because in all what we have been talking about, the object in question is just subject to a speed v, so if we change our system of coordinates (x, y, z) such that x is parallel to v, all the previous 1D results from chapter 3 will apply, with y, z being irrelevant.

In fact, for most of the formulae to be extended, considering the 2D case is enough, because adding quantities like speeds or momenta in 3D is in fact a 2D problem, and the extra irrelevant dimension is often something cumbersome. On the other hand, as seen in chapters 1-2, it is often better to switch directly to $\mathbb{R}^3$, as to take advantage of the magic of the vector product $\times$, which is defined only there. We will follow this latter path.

Let us start our discussion with a look at the non-relativistic case. Assuming that the object moves with speed v in the x direction, the frame change is given by:

$$x' = x - vt$$

$$y' = y$$

$$z' = z$$

$$t' = t$$

To be more precise, here the first 3 equations come from the law of motion, and $t' = t$ is the old $t' = t$. In the relativistic setting now, the result is more tricky, as follows:

THEOREM 4.3. *In the context of a relativistic object moving with speed v along the x axis, the frame change is given by the Lorentz transformation*

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

$$y' = y$$

$$z' = z$$

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

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

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

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

Regarding y, z, these are obviously unchanged, so done with these too. Finally, regarding time t, a naive thought would suggest that this is subject to a Lorentz contraction by $1/\gamma$, but this is not true, and more thinking leads to the conclusion that we must use the reverse Lorentz transformation, given by the following formulae:

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

$$y = y'$$

$$z = z'$$

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

$$t' = \frac{x - \gamma x'}{\gamma v} = \frac{x - \gamma^2(x - vt)}{\gamma v} = \frac{\gamma^2 vt + (1 - \gamma^2)x}{\gamma v}$$

On the other hand, we have the following computation:

$$\gamma^2 = \frac{c^2}{c^2 - v^2} \implies \gamma^2(c^2 - v^2) = c^2 \implies (\gamma^2 - 1)c^2 = \gamma^2 v^2$$

Thus we can finish the computation of t' as follows:

$$t' = \frac{\gamma^2 vt + (1 - \gamma^2)x}{\gamma v} = \frac{\gamma^2 vt - \gamma^2 v^2 x/c^2}{\gamma v} = \gamma\left(t - \frac{vx}{c^2}\right)$$

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

4b. Linear algebra

The Lorentz transformation being linear, time to do some math. Since y, z are irrelevant, we will put them at the end, and put the time t first, as to be close to x. By multiplying as well the time equation by c, our system looks better, as follows:

$$ct' = \gamma(ct - vx/c)$$

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

$$y' = y$$

$$z' = z$$

In linear algebra terms, the result is as follows:

THEOREM 4.4. *The Lorentz transformation is given by*

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

where $\gamma = 1/\sqrt{1-v^2/c^2}$ *as usual, and where* $\beta = v/c$.

PROOF. In terms of $\beta = v/c$, replacing v, the system looks as follows:

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

But this gives the formula in the statement. □

As an illustration, let us verify that the inverse Lorentz transformation is indeed given by reversing the speed, $v \to -v$. With notations as in Theorem 4.3, the result is:

PROPOSITION 4.5. *The inverse of the Lorentz transformation is given by* $v \to -v$,

$$x = \gamma(x' + vt')$$
$$y = y'$$
$$z = z'$$
$$t = \gamma(t' + vx'/c^2)$$

where $\gamma = 1/\sqrt{1-v^2/c^2}$ *is as usual the Lorentz factor, identical for* v *and* $-v$.

PROOF. In terms of the formalism in Theorem 4.4, reversing the speed $v \to -v$ amounts in reversing the $\beta = v/c$ parameter there:

$$\beta \to -\beta$$

What we have to prove, in order to establish the result, is that by doing so, we obtain the inverse of the matrix appearing there, namely:

$$L = \begin{pmatrix}\gamma&-\beta\gamma&0&0\\-\beta\gamma&\gamma&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

But here we can restrict attention to the upper left corner, where we have, as desired:

$$\begin{pmatrix}\gamma&-\beta\gamma\\-\beta\gamma&\gamma\end{pmatrix}\begin{pmatrix}\gamma&\beta\gamma\\\beta\gamma&\gamma\end{pmatrix}=\begin{pmatrix}\gamma^2(1-\beta^2)&0\\0&\gamma^2(1-\beta^2)\end{pmatrix}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

Now by getting back to notations as in Theorem 4.3, we obtain the result. □

There are many other things that can be said about all this, at the theoretical level. We will be back to the Lorentz transformation later, in the theoretical context, when discussing the Maxwell equations. In fact, as previously mentioned, the Lorentz transformation was found by Lorentz himself, before Einstein, working on electromagnetism.

4c. Speeds, revised

As a first application of the Lorentz transformation, we can now find a better explanation for the addition formula for speeds, that we know from chapter 2:

THEOREM 4.6. *The speed addition formula in 3D relativity is*

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

in $c = 1$ units.

PROOF. This formula, that took us the whole chapter 2 to establish, based on nothing or almost, is indeed quite easy to establish now, by using the power and magic of the Lorentz transformation, simply by computing some derivatives. □

Going ahead with the demolition of what we already know, we have as well:

THEOREM 4.7. *The magnitude of summed speeds in 3D relativity is*

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

as usual in $c = 1$ units.

PROOF. This is again something that we seriously struggled with, and which follows now quite easily, by using the magic of the Lorentz transformation. □

Very nice all this, you must agree, and with the comment that, with the Lorentz transformation trivializing everything, why not having started our book with it, as so many authors do. Well, the problem with the Lorentz transformation is that this is based on Principle 4.1, and go believe in that principle, without due preparations.

But probably best here is to ask for opinions, and we will ask the cat, the dog and the mouse, what they think about all this. Cat first, who is full of wisdom, as usual:

CAT 4.8. *I learned special and general relativity as a kitten, with speed addition coming via Lorentz. But if dog and mouse need to learn too, sure, go slow.*

Not bad, although, and this remains between us, I find cat to be sometimes a bit annoying. Dog on his side seems fully supporting:

DOG 4.9. *I learned special relativity as a freshman, also via Lorentz, and always thought of that as some kind of weird stuff. Good to have it slowly explained.*

Very nice, we are on the same wavelength, and no wonder here, since we share the same diet, and core values. But let's see now what mouse has to say:

MOUSE 4.10. *I find this treatment of relativity disrespectful towards us mice, who don't understand anything anyway, and would have preferred to have speed addition quickly explained via Lorentz. I intend to complain about this.*

Well, can't make everyone happy. This being said, what I find interesting in what mouse says is that he agrees with cat on the treatment of relativity. Starting from tonight they will share the same room, for discussing relativity before going to sleep.

4d. Some geometry

Some geometry.

4e. Exercises

Exercises.

Part II

Curved spacetime

Get up, stand up
Stand up for your right
Get up, stand up
Don't give up the fight

CHAPTER 5

Curved spacetime

5a. Spacetime separation

We have seen so far that basic classical mechanics, meaning the laws of motion, lead to a contradiction when it comes to light. However, everything can be fixed by using Einstein's theory of special relativity, whose main principle is that time and space are related, and curved. In the remainder of the present book we discuss how relativity, and curved spacetime, fit with the other known physics. We will do this in two steps:

(1) In the present Part II we further discuss the relativistic spacetime, and then we go into the study of high-speed particles, also called relativistic, including the photon which carries light, which appear from electromagnetism and quantum mechanics.

(2) With this done, the remaining piece of physics still needing to be unified with special relativity will be advanced classical mechanics, meaning the theory of gravity. We will discuss this in Part III, with applications to cosmology discussed in Part IV.

As a comment here, you might wonder why not doing this the other way around, first by talking about gravity, which is something well understood, and then leaving electromagnetism and quantum mechanics for later. Good point, but the thing is, as we will soon discover, electromagnetism and quantum mechanics, as developed by Maxwell, Lorentz and others, are already relativistic, so our "unification" work here will be easy task, basically reviewing the known theory, with relativistic ideas in mind. On the opposite, in what concerns classical gravity, as developed by Newton and others, this is definitely not relativistic, and making it relativistic will turn to be a quite tricky business.

Getting started now with the present Part II, curved spacetime in connection with non-gravitational physics, we need a plan, and goals. Here are some good questions:

Questions 5.1. *What does curved spacetime tell us about:*

(1) *Life and philosophy.*
(2) *Elementary particles.*
(3) *Mass and energy.*
(4) *Radiation and light.*

We will discuss these questions in the present chapter 5, and then in chapters 6,7,8. With two remarks however, first coming the fact that the present chapter 5, dealing with general spacetime abstractions, will be less philosophical than it seems, because what we will be doing here will be very useful when talking later cosmology, in Part IV of this book. And then coming the fact that chapter 7 will contain among others a formula that you surely would like to know more about, and as soon as possible, namely:

$$E = mc^2$$

In short, we have a good plan, with perhaps things which are a bit abstract for the present chapter 5, and for the next chapter 6 as well. But this is how the theory is best developed, and for hot stuff like $E = mc^2$, no worries, that is not long from now.

Getting started for good now, let us try to understand how space $\mathbb{R}^3$ and time $\mathbb{R}$ are exactly intricated. We already worked on this in chapter 4, with lots of formulae for the Lorentz transform. So, to start with, let us recall from there that we have:

FACT 5.2. *In the context of a relativistic object moving with speed v along the x axis, the frame change is given by the Lorentz transformation*

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

$$y' = y$$

$$z' = z$$

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

with $\gamma = 1/\sqrt{1 - v^2/c^2}$. Equivalently, the Lorentz transformation is given by

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

in matrix form, where $\beta = v/c$.

We refer to chapter 4 for more on all this, story and details, and with the comment that this is in fact more of a Theorem than a Fact. As another comment, the above two formulations of the Lorentz transform are of course equivalent, and the first one, with variables (x, y, z, t), certainly looks more familiar, but in practice we will rather tend to use the second one, with variables (ct, x, y, z), due to obvious linear algebra reasons. To be more precise, when a matrix is block-diagonal that is an extremely good thing, so choose your basis such as your matrix to be obviously block-diagonal in that basis.

Getting to work now, in non-relativistic physics two events are separated by space Δx and time Δt, with these two separation variables being independent. In relativistic physics this is no longer true, and the correct analogue of this comes from:

THEOREM 5.3. *The following quantity, called relativistic spacetime separation*

$$\Delta s^2 = c^2\Delta t^2 - (\Delta x^2 + \Delta y^2 + \Delta z^2)$$

is invariant under relativistic frame changes.

PROOF. We must prove that the quantity $K = c^2t^2 - x^2 - y^2 - z^2$ is invariant under Lorentz transformations. For this purpose, observe that we have:

$$K = \left\langle \begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}\begin{pmatrix}ct\\x\\y\\z\end{pmatrix}, \begin{pmatrix}ct\\x\\y\\z\end{pmatrix}\right\rangle$$

Now recall from Fact 5.2 that the Lorentz transformation is given by the following formula, where $\gamma = 1/\sqrt{1 - v^2/c^2}$ as usual, and where $\beta = v/c$:

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

Thus, if we denote by L the matrix of the Lorentz transformation, and by E the matrix found before, we must prove that for any vector ξ we have:

$$< E\xi, \xi >=< EL\xi, L\xi >$$

Since L is symmetric we have $< EL\xi, L\xi >=< LEL\xi, \xi >$, so we must prove:

$$E = LEL$$

But this is the same as proving $L^{-1}E = EL$, and by using the fact that $L \to L^{-1}$ is given by $\beta \to -\beta$, what we eventually want to prove is that:

$$L_{-\beta}E = EL_\beta$$

So, let us prove this. As usual we can restrict the attention to the upper left corner, call that NW corner, and here we have the following computations:

$$(L_{-\beta}E)_{NW} = \begin{pmatrix}\gamma&\beta\gamma\\\beta\gamma&\gamma\end{pmatrix}\begin{pmatrix}1&0\\0&-1\end{pmatrix} = \begin{pmatrix}\gamma&-\beta\gamma\\\beta\gamma&-\gamma\end{pmatrix}$$

$$(EL_\beta)_{NW} = \begin{pmatrix}1&0\\0&-1\end{pmatrix}\begin{pmatrix}\gamma&-\beta\gamma\\-\beta\gamma&\gamma\end{pmatrix} = \begin{pmatrix}\gamma&-\beta\gamma\\\beta\gamma&-\gamma\end{pmatrix}$$

The matrices on the right being equal, this gives the result. □

5b. More mathematics

More mathematics.

5c. Further physics

Further physics.

5d. Some conclusions

Some conclusions.

5e. Exercises

Exercises.

CHAPTER 6

Relativistic particles

6a. Charges, Gauss

In this chapter and in the next two ones we discuss the relation between relativity theory and electromagnetism, and with electromagnetism actually spilling from time to time into quantum mechanics. Our goals will be rather theoretical, but at the same time practical as well, and as a concrete list of questions to be solved, we have:

QUESTIONS 6.1. *Regarding the high-speed particles, also called relativistic:*

(1) *Are the electrons and other atomic constituents relativistic?*
(2) *Can we derive something macroscopic out of this?*
(3) *What about light, can it be included in the theory?*

We will see, in the present chapter 6, following Maxwell, Lorentz and others, that the answer to question (1) is yes. Based on this, we will study question (2) in chapter 7, following Einstein, and we will come up with a frightening formula there, $E = mc^2$, along with some illustrations for it, coming from stars and atomic bombs. Finally, in chapter 8 we will discuss question (3), again with a yes answer, and with, as a bonus, following Planck and others, the fact that heat can be included as well in the theory.

Getting started now, 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 6.2. *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 6.2 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 6.3. *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 graviation 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 6.2. Thus, we need void.

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

Fact 6.4 (Coulomb law). *Any pair of charges $q_1, q_2 \in \mathbb{R}$ is subject to a force as follows, which is attractive if $q_1q_2 < 0$ and repulsive if $q_1q_2 > 0$,*

$$||F|| = K \cdot \frac{|q_1q_2|}{d^2}$$

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 $x_1 \to x_2$ 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 6.5. *The Coulomb force of q_1 at x_1 acting on q_2 at x_2 is*

$$F = K \cdot \frac{q_1q_2(x_2 - x_1)}{||x_2 - x_1||^3}$$

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

Proof. We have indeed the following computation:

$$\begin{aligned} F &= sgn(q_1q_2) \cdot ||F|| \cdot \frac{x_2 - x_1}{||x_2 - x_1||} \\ &= sgn(q_1q_2) \cdot K \cdot \frac{|q_1q_2|}{||x_2 - x_1||^2} \cdot \frac{x_2 - x_1}{||x_2 - x_1||} \\ &= K \cdot \frac{q_1q_2(x_2 - x_1)}{||x_2 - x_1||^3} \end{aligned}$$

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

In relation now with the value of the constant K appearing above, we have:

Fact 6.6. *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.

Let us develop now the basic math needed for electrostatics. We first have:

Definition 6.7. *Given charges $q_1, \ldots, q_k \in \mathbb{R}$ located at positions $x_1, \ldots, x_k \in \mathbb{R}^3$, 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 $F = QE$.

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 6.8. *The electric field of a charge configuration $L\subset\mathbb R^3$, with charge density function $\rho:L\to\mathbb R$, 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 $F=QE$.

It is most convenient now to forget about the charges, and focus on the corresponding electric fields E. These fields are by definition vector functions $E:\mathbb R^3\to\mathbb R^3$, 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, constructed as follows:

Definition 6.9. *The field lines of an electric field $E:\mathbb R^3\to\mathbb R^3$ 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:

$$\begin{matrix}\nwarrow&\uparrow&\nearrow\\ \leftarrow&\oplus&\rightarrow\\ \swarrow&\downarrow&\searrow\end{matrix}\qquad\qquad\begin{matrix}\searrow&\downarrow&\swarrow\\ \rightarrow&\ominus&\leftarrow\\ \nearrow&\uparrow&\nwarrow\end{matrix}$$

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:

$$\begin{matrix}\nearrow&\nearrow&\rightarrow&\rightarrow&\rightarrow&\rightarrow&\searrow&\searrow\\ \nwarrow&\uparrow&\nearrow&\rightarrow&\rightarrow&\searrow&\downarrow&\swarrow\\ \leftarrow&\oplus&\rightarrow&\rightarrow&\rightarrow&\rightarrow&\ominus&\leftarrow\\ \swarrow&\downarrow&\searrow&\rightarrow&\rightarrow&\nearrow&\uparrow&\nwarrow\\ \searrow&\searrow&\rightarrow&\rightarrow&\rightarrow&\rightarrow&\nearrow&\nearrow\end{matrix}$$

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:

$$\begin{matrix}\leftarrow&\nwarrow&\nwarrow\\ &\uparrow&\nearrow\\ \leftarrow&\oplus&\\ &\downarrow&\searrow\\ \leftarrow&\swarrow&\swarrow\end{matrix}\qquad\qquad\begin{matrix}\nearrow&\nearrow&\rightarrow\\ \nwarrow&\uparrow&\\ &\oplus&\rightarrow\\ \swarrow&\downarrow&\\ \searrow&\searrow&\rightarrow\end{matrix}$$

The field lines 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 $x \in \mathbb{R}$ 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 6.10. *Given an electric field $E : \mathbb{R}^3 \to \mathbb{R}^3$, the knowledge of its field lines is the same as the knowledge of the composition*

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

where $S \subset \mathbb{R}^3$ is the unit sphere, and $n : \mathbb{R}^3 \to S$ 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. □

Let us introduce now a key definition, as follows:

Definition 6.11. *The flux of an electric field $E : \mathbb{R}^3 \to \mathbb{R}^3$ through a surface $S \subset \mathbb{R}^3$, assumed to be oriented, is the quantity*

$$\Phi_E(S) = \int_S < E(x), n(x) > dx$$

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.

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

PROPOSITION 6.12. *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:

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

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

More generally now, we have the following result:

THEOREM 6.13. *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:

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

(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:

$$\begin{aligned}
\Phi_E(S) &= Kq\int_S \frac{1-rx}{(1-2xr+r^2)^{3/2}}\, du\\
&= Kq\int_0^{2\pi}\int_0^{\pi} \frac{1-r\cos s}{(1-2r\cos s+r^2)^{3/2}}\cdot \sin s\, ds\, dt\\
&= 2\pi Kq\int_0^{\pi}\frac{(1-r\cos s)\sin s}{(1-2r\cos s+r^2)^{3/2}}\, ds\\
&= \frac{q}{2\varepsilon_0}\int_0^{\pi}\frac{(1-r\cos s)\sin s}{(1-2r\cos s+r^2)^{3/2}}\, ds
\end{aligned}$$

(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 $0, q/\varepsilon_0$ values of the flux.

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 6.14 (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 6.13, or even from Proposition 6.12, 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 [33], 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 6.13, by assuming this time that the integration takes place on an arbitrary surface as follows:

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

To be more precise, here $\lambda : S \to (0, \infty)$ 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 $x \to u$, and in the end, when integrating over S as in the proof of Theorem 6.13, this function λ dissapears.

(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, a bit later.

(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 [41]. 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 summmarize, 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 6.12 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 $q_1, \ldots, q_k \in \mathbb{R}$, and an exterior total charge Q_{ext}. We can surround each of $q_1, \ldots, q_k$ by small disjoint spheres $U_1, \ldots, U_k$,

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 6.12. □

We have the following point of view on the Gauss formula, more conceptual:

Theorem 6.15 (Gauss). *Given an electric potential E, its divergence is given by*

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

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:

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

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:

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

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:

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

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. □

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 6.16. *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 dissapeared, because in the gravitational context $m_1m_2>0$, always true, corresponds to an attractive force, while in our setting $q_1q_2>0$ 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. □

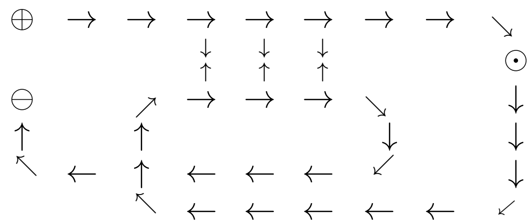
There are of course many other things that can be said, about electrostatics.

6b. Currents, Faraday

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:

$$\begin{matrix}
\oplus & \to & \to & \to & \to & \to & \to & \to & \searrow \\
 & & & \uparrow & \uparrow & \uparrow & & & \\
 & & & \downarrow & \downarrow & \downarrow & & & \odot \\
\ominus & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \swarrow
\end{matrix}$$

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:

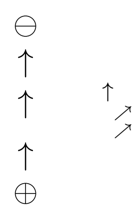


Summarizing, just by feeding a light bulb with a battery, and looking at the cables, and playing a bit with them, we are led to the following interesting conclusion:

FACT 6.17. *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 6.18 (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 6.17. 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 6.19. *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 6.19.

(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 6.18, 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 6.20 (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 6.18, and in the definition of $\times$. In fact, the Lorentz force law is just a fancy mathematical reformulation of Fact 6.18, 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 6.18.

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

THEOREM 6.21. *Magnetic forces do not work.*

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

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

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

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 6.22. *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 6.20 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 6.23. *The volume current density J satisfies*

$$< \nabla, J >= -\dot{\rho}$$

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 chage conservation:

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

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

Moving ahead now, let us formulate the following definition:

DEFINITION 6.24. *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:

$$< \nabla, J >= 0$$

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 6.25 (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 6.26 (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 previous discussion on this topic. In what regards the units, we won't be impressed by the ampere, and keep using the coulomb, as a main unit:

CONVENTIONS 6.27. *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 6.26, 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 6.28. *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:

$$\varepsilon_0\mu_0c^2 = 8.854 \times 1.256 \times 2.997^2 \times 10^{16-12-6} = 0.998$$

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

THEOREM 6.29. *We have the following formula:*

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

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:

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

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

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

THEOREM 6.30 (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:

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

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

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

THEOREM 6.31. *The equations of magnetostatics are*

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

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

with the second equation being the Ampère law.

PROOF. This follows indeed from the above discussion, and more specifically from Theorem 6.29 and Theorem 6.30, which both follow from the Biot-Savart law. □

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

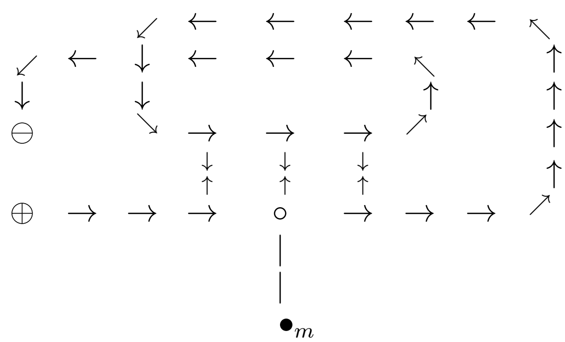
$$< \nabla, E >= \frac{\rho}{\varepsilon_0}$$

$$\nabla \times E = 0$$

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 6.17, but unfortunately, when doing the engineering and building such a motor, we are led to:

THEOREM 6.32. *A basic electomagnetic 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 6.21, 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 electomagnetic motors can be built, by tricking the 3D in Theorem 6.21, the idea being that of replacing straight wires by coils. □

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 6.33 (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:

$$\begin{array}{ccc}
\text{Silver} & : & 1.59 \times 10^{-8} \\
\text{Iron} & : & 9.61 \times 10^{-8} \\
\text{Graphite} & : & 1.6 \times 10^{-5} \\
 & - & \\
\text{Seawater} & : & 0.2 \\
\text{Diamond} & : & 2.7 \\
\text{Silicon} & : & 2500 \\
 & - & \\
\text{Water} & : & 8300 \\
\text{Glass} & : & 10^9 - 10^{14} \\
\text{Teflon} & : & 10^{22} - 10^{24}
\end{array}$$

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

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

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 6.35 (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:

$$\begin{matrix}
\oplus & \to & \to & \to & \to & \to & \to & \to & \searrow & \\
\star & & & & & & & & & \odot \\
\ominus & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \leftarrow & \swarrow &
\end{matrix}$$

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 6.36. *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

$$\mathcal{E} = \int_\gamma < F(x), dx >$$

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:

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

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 6.35, these can be fine-tuned, and reformulated in terms of the emf, in the following way:

FACT 6.37 (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:

$$\Phi = \int_\gamma < B(x), dx >$$

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

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

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 6.38. *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 6.32. To be more precise, that attempt failed due to the math in Theorem 6.21, telling us that magnetic forces do not work. But now we know from Fact 6.37 how to trick Theorem 6.21, by replacing straight wires by loops, or even bettter, 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 6.37. □

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 [41]. And with the remark however that, for serious applications, you are in need afterwards of a solid engineering book, centered on electromechanics.

6c. Maxwell equations

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

FACT 6.39 (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 6.40 (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 doublecheking 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 6.41 (Maxwell). *Electrodynamics is governed by the formulae*

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

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

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

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

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.

(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. □

So long for the Maxwell equations. We are now into deep physics, but as a comment, these are not the end of everything. The point indeed is that in the context of the simplest 2-body problem of electrodynamics, namely an electron moving around a proton, which corresponds to how the hydrogen atom works, the Maxwell equations do not apply well, and lead to a contradiction, and must be replaced by quantum mechanics. This reminds a bit the story of classical mechanics, with Newton being replaced by Einstein, but in the present case, interestingly, the bug is not of relativistic issue. More on this next.

6d. 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 6.42. *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:

$$E'_x = E_x$$
$$E'_y = \gamma(E_y - vB_z)$$
$$E'_z = \gamma(E_z + vB_y)$$

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

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

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

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

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

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

(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. □

There are many other things that can be said here, in relation with Einstein's relativity. We will come back to this later in this book, on numerous occasions.

6e. Exercises

Exercises.

CHAPTER 7

Mass and energy

7a. Collisions, momentum

We have learned many interesting things in the previous chapter, with the main achievement being that of having now a full menagery of fast-moving particles, obeying to relativity, which are complementary to the good old light that we have been using since chapter 1. Moreover, these new particles are in fact far more adapted to experimental physics than the ever-slippery light, because they are a bit similar to the macroscopic particles, having mass, momentum, energy which are far easier to measure.

So, many things to be done, as a continuation of all this, and as a collection of questions that we can potentially solve, if theory and experiments go well, we have:

Questions 7.1. *In the context of particles, microscopic and then macroscopic:*

(1) *What can we say about momentum p?*
(2) *What can we say about mass m?*
(3) *What can we say about energy E?*

We will see, following Einstein [**28**], that these questions are extremely interesting, leading to all sorts of bizarre new formulae and concepts and physics, including the formula $E=mc^2$, that you surely have heard about. And, as a supplementary piece of advertisement, this latter formula, $E=mc^2$, will be in fact so strange that, despite us being sure of what we're doing with relativity, and this formula coming as a theorem, we will have to scratch our heads, and look for experimental verifications of it. And with these verifications being fascinating, apocalyptic matter. But more on this later.

Getting started now, in order to answer Question 7.1 (1), the idea will be very simple, namely that of colliding relativistic particles, and computing various momenta p. So, let us first review the usual collision theory, from classical mechanics. We first have:

Fact 7.2. *In the context of general linear motion, in the case of a collision between two bodies, m_1, m_2 travelling at speeds v_1, v_2, the total momentum of the system*

$$p = m_1 v_1 + m_2 v_2$$

is conserved. The same happens of course without collision either, and also for systems of N bodies, with $N \in \mathbb{N}$ arbitrary, with all sorts of collisions allowed between them.

As a first comment, is this really physics, or just some abstraction? We know that gravity is everywhere, and that the very existence of m_1, m_2 leads to their gravity, and so to the negation of the general linear motion setting above. However, two trains colliding is certainly physics, and even scary physics, and this has nothing to do with gravity. Thus, what we have here is a true physics principle, dealing with real-life situations.

In order to understand now what is going on, consider two objects as in Fact 7.2, travelling first in 1D, for simplifying, and bound for collision:

$$\circ_{m_1} \to_{v_1} \qquad\qquad \leftarrow_{v_2} \circ_{m_2}$$

We know from real life that two things can happen, in this situation. The first case is that of an inelastic, also called plastic collision, where m_1, m_2 decide when meeting that they love each other, and pursue their journey as a couple, $m = m_1 + m_2$:

$$\bullet_m \to_v$$

Of course, who really knows what really happens during a plastic collision, at the microscopic level, but assuming somehow that no energy or something is dissipated, during that hot encounter, Fact 7.2 holds indeed, and allows us to do the math.

To be more precise, the math is quite simple, so let us upgrade right away our discussion, to the case where we have two bodies colliding, in arbitrary N dimensions:

$$\begin{matrix} \circ_{m_1} & & & \circ_{m_2} \\ & \searrow_{v_1} & \swarrow_{v_2} & \end{matrix}$$

As a result of our collision, we have a new body $m = m_1 + m_2$, with speed v:

$$\begin{matrix} \bullet_m \\ \downarrow_v \end{matrix}$$

The math, coming from the conservation of momentum, is very simple, as follows:

PROPOSITION 7.3. *In the context of a plastic collision between two bodies,*

$$m = m_1 + m_2 \quad , \quad v = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2}$$

are the mass and speed of the resulting body.

PROOF. This follows straight from Fact 7.2, because the momentum of $m = m_1 + m_2$ equals the sum of the initial momenta of m_1, m_2, and is therefore given by:

$$mv = m_1 v_1 + m_2 v_2$$

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

The second case now, that can happen as well, is that of an elastic collision. This is something more complicated, so let us go back to 1D, the situation being:

$$\circ_{m_1} \to_{v_1} \qquad\qquad \leftarrow_{v_2} \circ_{m_2}$$

The elastic collision is then the case opposed to love, with our two bodies meeting, comparing their m_i, v_i, then exchanging some speed depending on that, via a few quick fists, and then either keeping travelling forward, but slower, or going backwards:

$$\bullet_{m_1} \to_{v_1'} \qquad\qquad \circ_{m_2} \to_{v_2'}$$

$$\leftarrow_{v_1'} \circ_{m_1} \qquad\qquad \leftarrow_{v_2'} \bullet_{m_2}$$

$$\leftarrow_{v_1'} \circ_{m_1} \qquad\qquad \circ_{m_2} \to_{v_2'}$$

In the above pictures, the winner, which was m_1 in the first case, and m_2 in the second case, was awarded a black belt. As for the third case, that is some sort of draw.

Getting back now to the conservation of momentum, from Fact 7.2, it is pretty much clear that what we have there won't allow us to do the math. To be more precise, we can get from there only 1 equation, which is not enough for computing the output data. Fortunately, in the case of elastic collisions, Fact 7.2 can be complemented with:

FACT 7.4. *In the context of general linear motion, in the case of an elastic collision between two bodies, m_1, m_2 travelling at speeds v_1, v_2, the total energy of the system*

$$E = \frac{m_1||v_1||^2}{2} + \frac{m_2||v_2||^2}{2}$$

is conserved. The same happens of course without collision either, and also for systems of N bodies, with $N \in \mathbb{N}$ arbitrary, with multi-elastic collisions allowed between them.

Again, as in the case of the plastic collisions, who really knows what really happens during an elastic collision, at the microscopic level, but again, assuming that no things are lost, during that event, Fact 7.4 holds indeed, and allows us to do the math.

As another comment, while the formula of the momentum $p = mv$ from Fact 7.2 was something quite simple and intuitive, the above formula of the energy $E = m||v||^2/2$ is obviously something more subtle. We will be back to this, later.

Going ahead now, let us first investigate, just out of curiosity, what happens to the energy during a plastic collision. The result here, contradicting our previous guess that the moment conservation comes somehow from "no energy lost", is as follows:

THEOREM 7.5. *In the context of a plastic collision between two bodies, we have:*

$$E < E_1 + E_2$$

That is, some of the initial energy gets dissipated during the collision.

Proof. We use the equations found in Proposition 7.3, namely:

$$m = m_1 + m_2 \quad , \quad v = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2}$$

According to our definition of energy, from Fact 7.4, the initial energy is:

$$E_1 + E_2 = \frac{m_1 ||v_1||^2 + m_2 ||v_2||^2}{2}$$

As for the final energy, this is given by the following formula:

$$E = \frac{m||v||^2}{2} = \frac{||m_1 v_1 + m_2 v_2||^2}{2(m_1 + m_2)}$$

So, let us compute now the difference between these two quantities. We obtain:

$$\begin{aligned} E_1 + E_2 - E &= \frac{(m_1 + m_2)(m_1 ||v_1||^2 + m_2 ||v_2||^2) - ||m_1 v_1 + m_2 v_2||^2}{2(m_1 + m_2)} \\ &= \frac{m_1 m_2 (||v_1||^2 + ||v_2||^2) - 2 < m_1 v_1, m_2 v_2 >}{2(m_1 + m_2)} \\ &= \frac{m_1 m_2 (||v_1||^2 + ||v_2||^2 - 2 < v_1, v_2 >)}{2(m_1 + m_2)} \end{aligned}$$

But the quantity on top right is subject to the following inequality, valid for any two vectors $v_1, v_2 \in \mathbb{R}^N$, and with the equality case happening precisely when $v_1 = v_2$:

$$||v_1||^2 + ||v_2||^2 \geq 2 < v_1, v_2 >$$

Thus $E_1 + E_2 \geq E$, and since a collision cannot happen when the inital speeds are the same, $v_1 = v_2$, the equality case cannot happen, and so $E_1 + E_2 > E$, as stated. □

As already mentioned, the above result might seem quite surprising, contradicting our previous guess that the moment conservation principle comes from something of type "no energy lost". Let us record this finding in the form of an informal statement:

Conclusion 7.6. *Momentum ain't the same thing as energy.*

Moving ahead now, and back to the elastic collisions, the two conservation principles that we have, namely Fact 7.2 and Fact 7.4, allow us to do the math. Let us first work out the case of an ellastic collision in 1D, the initial picture being as follows:

$$\circ_{m_1} \rightarrow_{v_1} \qquad\qquad \leftarrow_{v_2} \circ_{m_2}$$

Depending on the resulting fight, we can have either a win or m_1 or m_2, or a draw. Abstractly however, we can simply say that we are in a draw situation, the picture being as follows, with the convention that we do not know yet the directions of v_1', v_2':

$$\leftarrow_{v_1'} \circ_{m_1} \qquad\qquad \circ_{m_2} \rightarrow_{v_2'}$$

With these conventions made, the precise 1D result is as follows:

Proposition 7.7. *In the context of a* 1D *elastic collision between two bodies,*

$$v_1' = \frac{(m_1 - m_2)v_1 + 2m_2v_2}{m_1 + m_2}$$

$$v_2' = \frac{(m_2 - m_1)v_2 + 2m_1v_1}{m_1 + m_2}$$

are the resulting speeds of the two bodies.

Proof. According to our momentum and energy conservation principles from Fact 7.2 and Fact 7.4, the resulting speeds v_1', v_2' satisfy the following two equations:

$$m_1v_1 + m_2v_2 = m_1v_1' + m_2v_2'$$

$$m_1v_1^2 + m_2v_2^2 = m_1v_1'^2 + m_2v_2'^2$$

Now observe that these equations can be written as follows:

$$m_1(v_1 - v_1') = m_2(v_2' - v_2)$$

$$m_1(v_1^2 - v_1'^2) = m_2(v_2'^2 - v_2^2)$$

By dividing the second equation by the first one, our system becomes:

$$m_1(v_1 - v_1') = m_2(v_2' - v_2)$$

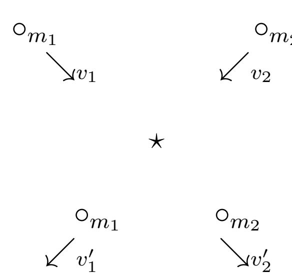
$$v_1 + v_1' = v_2' + v_2$$

And by doing now the math, we are led to the formulae in the statement. □

Getting now to arbitrary N dimensions, things here become more complicated, due to a number of reasons. As a first observation, which is actually very good news, we are in fact in 2 dimensions, and more specifically in the following plane:

$$E = span(v_1, v_2)$$

Thus we may assume $E = \mathbb{R}^2$, with the collision point being the origin $0 \in \mathbb{R}^2$, and do our computations here. Schematically, we can represent the collision as follows:



Let us quickly do the math. We have to compute two vectors v_1', v_2', accounting for a total of 4 numbers. But what we have is one vector equation, coming from momentum, and one scalar equation, coming from energy, so a total of 3 equations. Thus, we won't be able to do the math, unless we know something more on the collision mechanism.

To be more precise here, as already mentioned on numerous occasions, we don't really know what happens, at the microscopic level, when collisions take place, be them plastic or elastic, or of any other kind. And in the case of elastic collisions in $N \geq 2$ dimensions, there are several possible mechanisms, which in practice each correspond to a different way in which the original angles of v_1, v_2 get transformed into the output angles of v_1', v_2'. More specifically, the extra parameter that we need is some sort of "scattering angle" θ, coming from the precise mechanism of collision that we are investigating.

But this is something quite complicated, that we won't get into, at this stage. So long story short, and getting back now to math, in the absence of more data, and physical knowledge in general, our tools will be Fact 7.2 and Fact 7.4, and we will be able to solve the problem up to 1 degree of freedom. So, let us do this. The result is as follows:

Theorem 7.8. *In the context of an elastic collision between two bodies, assumed without loss of generality to happen in $\mathbb{R}^2$, the output speeds are*

$$v_1' = v_1 + \frac{q}{m_1} \quad , \quad v_2' = v_2 - \frac{q}{m_2}$$

with the vector parameter $q \in \mathbb{R}^2$ being subject to the following equation:

$$2 < v_2 - v_1, q >= \left(\frac{1}{m_1} + \frac{1}{m_2}\right) ||q||^2$$

Thus, the collision problem is solved, up to an angle $\theta \in \mathbb{R}$.

Proof. We can solve this by using Fact 7.2 and Fact 7.4, as follows:

(1) According to our general momentum and energy conservation principles, the output speeds v_1', v_2' are subject to the following two equations:

$$m_1 v_1 + m_2 v_2 = m_1 v_1' + m_2 v_2'$$

$$m_1 ||v_1||^2 + m_2 ||v_2||^2 = m_1 ||v_1'||^2 + m_2 ||v_2'||^2$$

Let us first look at the first equation. This equation can be written as follows:

$$m_1(v_1' - v_1) = m_2(v_2 - v_2')$$

Now if we call q this quantity, which is the individual change of momentum, from the perspective of m_1, and from the opposite perspective of m_2, we have:

$$v_1' = v_1 + \frac{q}{m_1} \quad , \quad v_2' = v_2 - \frac{q}{m_2}$$

(2) Now let us plug these values into the second equation above. We obtain:

$$m_1 ||v_1||^2 + m_2 ||v_2||^2 = m_1 \left|\left| v_1 + \frac{q}{m_1} \right|\right|^2 + m_2 \left|\left| v_2 - \frac{q}{m_2} \right|\right|^2$$

By expanding the scalar products and then simplifying, we obtain:

$$\frac{||q||^2}{m_1}+\frac{||q||^2}{m_2}+2<v_1,q>-2<v_2,q>=0$$

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

(3) As for the last assertion, this is something rather philosophical. To be more precise, we know that our parameter $q\in\mathbb{R}^2$ is subject to an equation of the following type:

$$<v,q>=\lambda||q||^2$$

Now if we denote by $\theta\in\mathbb{R}$ the oriented angle between v,q, this equation reads:

$$||v||\cos\theta=\lambda||q||$$

Thus we can recover $||q||$, and so also $q\in\mathbb{R}^2$ itself, out of this angle $\theta\in\mathbb{R}$, and this leads to the conclusion that our problem is indeed solved up to an angle $\theta\in\mathbb{R}$. $\square$

As an illustration for Theorem 7.8, let us work out what happens in the case of a 1D collison. Here we already know the answer, from Proposition 7.7, but wait for it. So, let us get back to our usual 1D scheme, with m_1,m_2 about to collide, on a line:

$$\circ_{m_1}\to_{v_1}\qquad\qquad\leftarrow_{v_2}\circ_{m_2}$$

In the context of Theorem 7.8, the fact that we are now in 1D simply tells us that the speed vectors $v_1,v_2\in\mathbb{R}^2$ are proportional, $v_1=\lambda v_2$. But this does not force in any way the vector $q\in\mathbb{R}^2$ there to be aligned with v_1,v_2, or if you prefer, the angle $\theta\in\mathbb{R}$ there to be 0. Thus, we are led to the peculiar conclusion that our general elastic collision formalism, leading to Theorem 7.8, theoretically allows escapes from 1D to 2D.

What to do? Modesty as usual, this is what we have, and it's after all not that bad. And for closing this 1D discussion, let us however formulate, as a consequence of our new knowledge from Theorem 7.8, a better formulation of Proposition 7.7, as follows:

THEOREM 7.9. *In the context of a* 1D *elastic collision between two bodies, staying as normal in* 1D*, the resulting speeds of the two bodies are*

$$v_1'=v_1+\frac{q}{m_1}\quad,\quad v_2'=v_2-\frac{q}{m_2}$$

where $q\in\mathbb{R}$ is the individual change of momentum, given by

$$\left(\frac{1}{m_1}+\frac{1}{m_2}\right)q=2(v_2-v_1)$$

from the perspective of m_1, and from the opposite perspective of m_2.

Proof. This follows either from Proposition 7.7, or from Theorem 7.8:

(1) From the perspective of Proposition 7.7, we have done some quick algebra there, without really knowing what we're doing, leading to the following formulae:

$$v_1' = \frac{(m_1 - m_2)v_1 + 2m_2v_2}{m_1 + m_2} \quad , \quad v_2' = \frac{(m_2 - m_1)v_2 + 2m_1v_1}{m_1 + m_2}$$

Now observe that these two formulae can be alternatively written as follows:

$$v_1' = v_1 + \frac{2m_2(v_2 - v_1)}{m_1 + m_2} \quad , \quad v_2' = v_2 + \frac{2m_1(v_1 - v_2)}{m_1 + m_2}$$

But this leads to the formulae in the statement, and to that conclusion about q.

(2) From the perspective of Theorem 7.8, we have indeed the formulae of v_1', v_2' in the statement, and with q being the change of momentum, as stated. Thus, it remains to establish the precise formula of q. In the context of Theorem 7.8, the equation is:

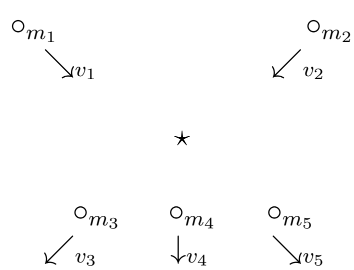
$$2 < v_2 - v_1, q >= \left(\frac{1}{m_1} + \frac{1}{m_2}\right) ||q||^2$$

Now since everything happens, by assumption, in 1D, the vectors v_1, v_2, q are pairwise proportional. Thus the above scalar product becomes a usual product, and also the squared norm becomes a usual square, and so our equation simply reads:

$$2(v_2 - v_1)q = \left(\frac{1}{m_1} + \frac{1}{m_2}\right) q^2$$

But this gives the formula of q in the statement, and we are done. □

Finally, again speaking generalities about collisions, rememeber that our first example was that of two trains colliding. But that type of collision, which is somewhat generic in real-life problems, is not elastic, neither plastic, but rather something in between, with the generic, 2D picture being something quite frightening, of the following type:



Again, this is something complicated, that we will not discuss at this stage. Let us point out, however, that the conservation of momentum, from Fact 7.2, does apply to such situations, and can be used in order to get information. More on this later.

Getting back now to relativity theory, we must discuss momentum, mass and energy, in the relativistic context. Things here are quite tricky, and as a first objective we would like to fix the momentum conservation equations for the plastic collisions, namely:

$$m = m_1 + m_2$$

$$mv = m_1 v_1 + m_2 v_2$$

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

FACT 7.10. *When defining the relativistic mass of an object of rest mass $m > 0$, moving at speed v, by the formula*

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

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

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

$$M = M_1 + M_2$$

$$Mv = M_1 v_1 + M_2 v_2$$

Summarizing, we have now a good relativistic theory of mass and momentum.

7b. Mass and energy

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

THEOREM 7.11. *The relativistic energy of an object of rest mass $m > 0$,*

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

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

$$E = mc^2$$

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

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

being called relativistic kinetic energy of m.

Proof. All this is a bit abstract, coming from Fact 7.10, as follows:

(1) Given an object of rest mass $m > 0$, consider its relativistic mass $M = \gamma m$, as appearing in Fact 7.10, and then consider the following quantity:

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

We know from Fact 7.10 that the relativistic mass M is conserved, so $\mathcal{E} = Mc^2$ is conserved too. In view of this, is makes somehow sense to call $\mathcal{E}$ energy. There is of course no clear reason for doing that, but let's just do it, and we'll understand later.

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

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

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

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

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

(3) But this leads to the conclusions in the statement. Indeed, we are certainly dealing with some sort of energies here, and so calling the above quantity $\mathcal{E}$ relativistic energy is legitimate, and calling $E = mc^2$ rest energy is legitimate too. Finally, the difference between these two energies $T = \mathcal{E} - E$ follows to be given by:

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

Thus, calling T relativistic kinetic energy is legitimate too, and we are done. □

As a conclusion to all this, we have now a full relativistic theory for the momentum p, the mass m, and the energy E. This is very nice, and completes the basics of Einstein's relativity theory, regarding the speed v, the time t, and the distance d, as developed in Part I of the present book. Of course, the last result, Theorem 7.11 dealing with energy, remains a bit unclear. Also, although all this is compatible with electromagnetism, there is still a mystery in what concerns gravity. We will be back to these questions.

7c. Atomic bombs

The formula $E = mc^2$ is quite frightening when thinking a bit numerics, to the point that you might now start fearing a calm glass of water, knowing how much energy is stored in it, and what will happen if that was to explode. Fortunately, water does not spontaneously explode, however other more specialized materials do.

So, let us go back first to the main energy formula that we have so far, $E = mc^2$, and try to understand its numerics, and consequences. At the terrestrial level, we have:

THEOREM 7.12. *An atomic bomb based on a glass of water releasing all its $E = mc^2$ energy is equivalent to the Giza Pyramid hitting you at 30 km/s.*

PROOF. When converting $E = mc^2$ into kinetic energy of a body M, the formula is:

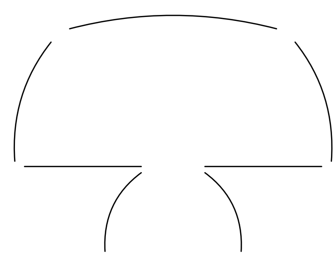
$$mc^2 = \frac{Mv^2}{2} \implies v = \sqrt{\frac{2m}{M}}\, c$$

In our case the glass of water m, glass included, is about 300 grams, and the Giza Pyramid M is about 6 million tons. Thus the impact speed is:

$$v = \sqrt{\frac{2 \times 3 \times 10^{-1}}{6 \times 10^9}}\, c = \frac{c}{10^5} = 30,000 \ m/s$$

As already mentioned, do not worry about this. Certain atomic bombs, based on other materials, can indeed be constructed. But the glass of water, as long as it stays far away from a big source of energy, of stellar or atomic type, will certainly not explode. □

In order to construct now atomic bombs, the normal method is that of using nuclear fission or fusion. This can be done even here on Earth, with spectacular results:



Tsar Bomba (CCCP)

The functioning of such bombs can be explained via quantum mechanics.

7d. Main sequence stars

However, atomic bombs are in fact far more familiar than this, because any star in the night sky is one, as strange as this might seem. There are many things that can be said here, and the general idea, at least from the viewpoint of gravity, is as follows:

FACT 7.13. *The stars fall into several classes, as follows:*

(1) *Main sequence stars, and their versions.*
(2) *Dwarfs, of various types.*
(3) *Neutron stars, of various types.*
(4) *Black holes, of various types.*

To be more precise here, (1) are the "usual" stars, whose functioning mechanisms, following Bethe and others, can be undestood via some quantum mechanics. Then (2) and (3) are also of quantum mechanical nature, and we will be back to all these objects later. As for (4), these are quite interesting for us, with our present knowledge. Indeed, while being of course real and observed, black holes are somehow purely mathematical objects, that can be studied by using the geometry tools from general relativity.

We will be back to stars, black holes and related topics later, in Part IV of the present book, after developing first, in Part III, the theory of general relativity.

Finally, as a last comment on all this, although the energy output of an atomic bomb exploding here on Earth can be explicitely measured, and leads to a solid confirmation of $E = mc^2$, measuring the energy output of the Sun, or of any other star, again with the formula $E = mc^2$ in mind, is a more complicated business. What we know about the energy of the Sun comes to us via its light, and so we will need to know more about light, and the energy that it carries. We will discuss this in the next chapter.

7e. Exercises

Exercises.

CHAPTER 8

Back to light

8a. Radiation

Remember light, that we started our book with? In this chapter we go back precisely to this, light. We have a whole palette of questions to be solved, as follows:

QUESTIONS 8.1. *Regarding light:*

(1) *What is light?*
(2) *Why is light colored?*
(3) *Why is deep red related to heat?*
(4) *Why do the Einstein principles hold?*

We will solve here all these questions, and with the solution of (4) bringing us to the Einstein principles regarding light and its speed $c < \infty$, that we started our book with. So, interesting theoretical loop, as we love them in physics, getting closed.

But perhaps not everyone in the audience is a physicist, loving theoretical loops, so time for a round-table discussion about this. As in chapter 4, where we dealt with other delicate pedagogical issues, we will ask the cat, the dog, and the mouse. Cat first:

CAT 8.2. *Advanced knowledge always lies on a loop, because of the Yin and the Yang, with the infinity above being the same as the infinity below.*

Thanks cat, this sounds quite deep, most likely requiring a good knowledge of both relativity and quantization, stuff well beyond my level. What about you, dog?

DOG 8.3. *I want to know about what's true and what's not, and whether that knowledge comes linearly, or on a loop, I don't mind much.*

Thanks dog, as usual when it comes to deep questions, we are on the same wavelength. As for mouse, I haven't seen him in a while, in fact since chapter 4, no idea what happened to him. But, there's some sort of whining that I can hear, probably him:

SPIRIT OF MOUSE 8.4. *Knowledge is like a tree, arising from a tremendous bureaucratic effort, and must be explained linearly. I intend to complain about this.*

Thanks mouse, I'm not that surprised, and in the hope that things going well for you, wherever you are, and away from too much heat. In any case, having you and your opinions around for a while has been, for myself, dog and cat, a delicious experience.

Back to work now, we must talk about light. Let us first recall:

FACT 8.5 (Maxwell). *Electrodynamics is governed by the formulae*

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

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

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

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

called Maxwell equations.

PROOF. This is something in between a fact and a theorem, with the formulae coming either from experiments, and from mathematics, that we know from chapter 6. □

We will need the following consequence of the Maxwell equations:

THEOREM 8.6. *In regions of space where there is no charge or current present the Maxwell equations for electrodynamics read*

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

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

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

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

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

PROOF. Under the circumstances in the statement, namely no charge or current present, the Maxwell equations simply read, taking into account $\mu_0 \varepsilon_0 = 1/c^2$:

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

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

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

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

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

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

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

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

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

So, what is light? Light is the wave predicted by Theorem 8.6, travelling in vacuum at the maximum possible speed, c, and with an important extra property being that it depends on a real positive parameter, that can be called, upon taste, frequency, wavelength, or color. And in what regards the creation of light, the mechanism here is as follows:

FACT 8.7. *An accelerating or decelerating charge produces electromagnetic radiation, called light, whose frequency and wavelength can be explicitely computed.*

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

Getting back now to Fact 8.7, in its general form, as stated above, this is something which can be deduced via some math, based on the Maxwell equations.

8b. Color, polarization

Moving ahead, let us go back to the wave equation $\ddot{\varphi} = v^2 \Delta\varphi$ from Theorem 8.6, and try to understand its simplest solutions. In 1D, the situation is as follows:

THEOREM 8.8. *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. □

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

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

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

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

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

Definition 8.9. *A monochromatic plane wave is a solution of the 3D wave equation which moves in only 1 direction, making it in practice a solution of the 1D wave equation, and which is of the special from found in Theorem 8.8, 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 8.8. And second is the assumption, in connection with the Fourier decomposition result from the end of Theorem 8.8, that our solution is of "pure" type, meaning a wave having a well-defined wavelenght and frequency, instead of being a "packet" of such pure waves.

All this is still mathematics, and making now the connection with physics and electromagnetism, and more specifically with Theorem 8.6 and Fact 8.7, we have:

Fact 8.10. *Physically speaking, a monochromatic plane wave is the electromagnetic radiation appearing as in Theorem 8.6 and Fact 8.7, via equations of type*

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

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

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

To be more precise, what we are doing here is to import the conclusions of our mathematical discussion so far, from Theorem 8.8 and Definition 8.9, into the context of our original physics discussion, from Fact 8.7. And also to add an extra twist coming from physics, and more specifically from the notion of polarization.

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

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

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

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

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

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

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

8c. Basic optics

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

FACT 8.11. *Inside a linear, homogeneous medium, where there is no free charge or current present, the Maxwell equations for electrodynamics read*

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

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

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

Observe that this is precisely the first part of Theorem 8.6, 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 8.6, we have:

THEOREM 8.12. *Inside a linear, homogeneous medium, where there is no free charge or free current present, both E and B are subject to the wave equation*

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

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

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

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

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

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

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

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 8.12 obviously deals with the transparent case.

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

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

Fact 8.13. *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 [**41**].

As a simple consequence of the above, we have:

Theorem 8.14. *Light can be decomposed, by using a prism.*

Proof. This follows from Fact 8.13. Indeed, when hitting a piece of glass, provided that the hitting angle is not 90°, the light will decompose over the wavelenghts present, with the corresponding refraction angles depending on these wavelengths. And we can capture these split components at the exit from the piece of glass, again deviated a bit, provided that the exit surface is not parallel to the entry surface. And the simplest device doing the job, that is, having two non-parallel faces, is a prism. □

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

Fact 8.15. *We can study events via spectroscopy, by capturing the light the event has produced, decomposing it with a prism, carefully recording its "spectral signature", consisting of the wavelenghts present, and their density, and then doing some reverse engineering, consisting in reconstructing the event out of its spectral signature.*

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

As a conclusion, we have learned many things about light, and in particular the method of spectroscopy, which is something very practical, and powerful. As an application now, which is probably the most famous, we can study the atoms by using this method.

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

Fact 8.16 (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 8.17 (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 8.18 (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 8.19 (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 $n_1 = 4, 5, 6, \ldots$, and these were discovered later, by Brackett in 1922, Pfund in 1924, Humphreys in 1953, and others aftwerwards, with all these extra lines being in far IR. The simplified complete table is as follows:

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

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

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

Thought 8.20. *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, we have some serious data here. The 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 8.21. *The Ritz-Rydberg combination principle reminds the formula*

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

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

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

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

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

Thought 8.22. *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.*

Time now to put all the pieces of the puzzle together. Conclusion 8.19, Thought 8.20 and Thought 8.21 are all we need, for developing quantum mechanics, and following Heisenberg, Schrödinger and many others, we are led to the following conclusion, along the lines initially suggested by Bohr, who was the initiator of the whole program:

Claim 8.23 (Bohr and others). *The atoms are formed by a core of protons and neutrons, surrounded by a cloud of electrons, basically obeying to a modified version of electromagnetism. And with a fine mechanism involved, as follows:*

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

And isn't this beautiful. Moreover, some further claims are that the theory can be further extended and fine-tuned as to explain many other phenomena, such as Einstein's work on the photoelectric effect, the work of Becquerel and Pierre and Marie Curie on radioactivity, and generally speaking, all the physics and chemistry known.

And the story is not over here. Following Heisenberg, the suppementary 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 8.21.

So long for light and quantum mechanics. For more on all this, you have plenty of good books, as for instance Feynman [34], Griffiths [42] or Weinberg [93]. By the way, do not forget to learn some basic chemistry too. Claim 8.23 is not the end of the story, but rather its beginning, the idea being that the atomic mechanism explained there allows the classification of the atoms into the Periodic Table, and also the understanding of the combination mechanisms between these atoms, producing molecules, and then life.

8d. Max Planck

Our discussion so far implicitely 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 8.24. *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:

$$E=Re(\widetilde{E})\quad:\quad\widetilde{E}=e_n\,e^{i(<k_n,x>-w_nt)}$$

$$B=Re(\widetilde{B})\quad:\quad\widetilde{B}=b_n\,e^{i(<k_n,x>-w_nt)}$$

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:

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

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

$$< k_n, e_n >=< k_n, b_n >= 0$$

(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):

$$E = Re(\widetilde{E}) \quad : \quad \widetilde{E} = \sum_n e_n \, e^{i(<k_n,x>-w_n t)}$$

$$B = Re(\widetilde{B}) \quad : \quad \widetilde{B} = \sum_n b_n \, e^{i(<k_n,x>-w_n t)}$$

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

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

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

$$\begin{aligned}
\frac{1}{V}\int_V ||E||^2 &= \frac{1}{4}\sum_n < e_n, e_n > + \frac{1}{4}\sum_n < e_n, \bar{e}_{-n} > e^{-2iw_nt} \\
&+ \frac{1}{4}\sum_n < \bar{e}_n, e_{-n} > e^{2iw_nt} + \frac{1}{4}\sum_n < \bar{e}_n, \bar{e}_n >
\end{aligned}$$

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

$$\begin{aligned}
\frac{1}{V}\int_V ||B||^2 &= \frac{1}{4}\sum_n < b_n, b_n > + \frac{1}{4}\sum_n < b_n, \bar{b}_{-n} > e^{-2iw_nt} \\
&+ \frac{1}{4}\sum_n < \bar{b}_n, b_{-n} > e^{2iw_nt} + \frac{1}{4}\sum_n < \bar{b}_n, \bar{b}_n >
\end{aligned}$$

(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:

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

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

$$< k_n, e_n > = < k_n, b_n > = 0$$

By using these formulae, we first obtain the following identity:

$$\begin{aligned} < b_n, b_n > &= \frac{c^2}{w_n^2} < k_n \times e_n, k_n \times e_n > \\ &= \frac{c^2||k_n||^2}{w_n^2} < e_n, e_n > \\ &= < e_n, e_n > \end{aligned}$$

Similarly, we have we well the following identity:

$$\begin{aligned} < b_n, \bar{b}_{-n} > &= \frac{c^2}{w_n^2} < k_n \times e_n, k_{-n} \times \bar{e}_n > \\ &= -\frac{c^2||k_n||^2}{w_n^2} < e_n, \bar{e}_{-n} > \\ &= - < e_n, \bar{e}_{-n} > \end{aligned}$$

Also similarly, we have as well the following identity:

$$\begin{aligned} < \bar{b}_n, b_{-n} > &= \frac{c^2}{w_n^2} < k_n \times \bar{e}_n, k_{-n} \times e_n > \\ &= -\frac{c^2||k_n||^2}{w_n^2} < \bar{e}_n, e_{-n} > \\ &= - < \bar{e}_n, e_{-n} > \end{aligned}$$

Finally, we have as well the following identity:

$$\begin{aligned} < \bar{b}_n, \bar{b}_n > &= \frac{c^2}{w_n^2} < k_n \times \bar{e}_n, k_n \times \bar{e}_n > \\ &= \frac{c^2||k_n||^2}{w_n^2} < \bar{e}_n, \bar{e}_n > \\ &= < \bar{e}_n, \bar{e}_n > \end{aligned}$$

(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:

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

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

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

(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, expect for the UV range, where things diverge. And with this phenomenon being called "UV catastrophe". □

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 result being as follows:

THEOREM 8.25. *The correct formula for the black body radiation, obtained by assuming that energy is quantized, is the Planck formula*

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

with h being a new constant, called Planck constant, given by

$$h=6.626\ 070\ 15\times10^{-34}$$

as per the latest SI regulations. The Planck formula fits with all known data, fits as well with the Rayleigh-Jeans formula 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:

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

PROOF. This is something quite technical, obtained along the lines of the proof of Theorem 8.24, by counting in a new way, by assuming that energy is quantized. □

Regarding applications, 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. We will be back to this in Part IV of the present book, when discussing cosmology.

8e. Exercises

Exercises.

Part III

General relativity

Gimme hope Joanna
Hope Joanna
Gimme hope Joanna
Before the morning come

CHAPTER 9

Some geometry

9a.

9b.

9c.

9d.

9e. Exercises

CHAPTER 10

General relativity

10a.

10b.

10c.

10d.

10e. Exercises

CHAPTER 11

Basic solutions

11a.

11b.

11c.

11d.

11e. Exercises

CHAPTER 12

Singularities

12a.

12b.

12c.

12d.

12e. Exercises

Part IV

Stars, black holes

One love, one heart
Let's get together and feel alright
One love, one heart
Let's praise to the Lord, and I will feel alright

CHAPTER 13

Cosmology

13a.

13b.

13c.

13d.

13e. Exercises

CHAPTER 14

Stars, fusion

14a.

14b.

14c.

14d.

14e. Exercises

CHAPTER 15

Compact stars

15a.

15b.

15c.

15d.

15e. Exercises

CHAPTER 16

Black holes

16a.

16b.

16c.

16d.

16e. Exercises

Bibliography

[1] V.I. Arnold, Ordinary differential equations, Springer (1973).

[2] V.I. Arnold, Mathematical methods of classical mechanics, Springer (1974).

[3] V.I. Arnold, Lectures on partial differential equations, Springer (1997).

[4] V.I. Arnold, Catastrophe theory, Springer (1974).

[5] V.I. Arnold and B.A. Khesin, Topological methods in hydrodynamics, Springer (1998).

[6] V.I. Arnold, V.V. Kozlov and A.I. Neishtadt, Mathematical aspects of classical and celestial mechanics, Springer (2006).

[7] N.W. Ashcroft and N.D. Mermin, Solid state physics, Saunders College Publ. (1976).

[8] M.F. Atiyah, The geometry and physics of knots, Cambridge Univ. Press (1990).

[9] T. Banica, Calculus and applications (2023).

[10] T. Banica, Introduction to modern physics (2023).

[11] T. Banica, Principles of thermodynamics (2023).

[12] G.K. Batchelor, An introduction to fluid dynamics, Cambridge Univ. Press. (1967).

[13] R.J. Baxter, Exactly solved models in statistical mechanics, Academic Press (1982).

[14] S.J. Blundell and K.M. Blundell, Concepts in thermal physics, Oxford Univ. Press (2006).

[15] S.M. Carroll, Spacetime and geometry, Cambridge Univ. Press (2004).

[16] P.M. Chaikin and T.C. Lubensky, Principles of condensed matter physics, Cambridge Univ. Press (1995).

[17] A.R. Choudhuri, Astrophysics for physicists, Cambridge Univ. Press (2012).

[18] D.D. Clayton, Principles of stellar evolution and nucleosynthesis, Univ. of Chicago Press (1968).

[19] A. Connes, Noncommutative geometry, Academic Press (1994).

[20] W.N. Cottingham and D.A. Greenwood, An introduction to the standard model of particle physics, Cambridge Univ. Press (2012).

[21] P.A. Davidson, Introduction to magnetohydrodynamics, Cambridge Univ. Press (2001).

[22] P. Di Francesco, Meander determinants, *Comm. Math. Phys.* **191** (1998), 543–583.

[23] P.A.M. Dirac, Principles of quantum mechanics, Oxford Univ. Press (1930).

[24] M.P. do Carmo, Differential geometry of curves and surfaces, Dover (1976).

[25] M.P. do Carmo, Riemannian geometry, Birkhäuser (1992).

[26] S. Dodelson, Modern cosmology, Academic Press (2003).

[27] R. Durrett, Probability: theory and examples, Cambridge Univ. Press (1990).

[28] A. Einstein, Relativity: the special and the general theory, Dover (1916).

[29] L.C. Evans, Partial differential equations, AMS (1998).

[30] W. Feller, An introduction to probability theory and its applications, Wiley (1950).

[31] E. Fermi, Thermodynamics, Dover (1937).

[32] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics I: mainly mechanics, radiation and heat, Caltech (1963).

[33] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics II: mainly electromagnetism and matter, Caltech (1964).

[34] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics III: quantum mechanics, Caltech (1966).

[35] A.P. French, Special relativity, Taylor and Francis (1968).

[36] S. Galtier, Introduction to modern magnetohydrodynamics, Cambridge Univ. Press (2016).

[37] H. Goldstein, C. Safko and J. Poole, Classical mechanics, Addison-Wesley (1980).

[38] D.L. Goodstein, States of matter, Dover (1975).

[39] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory, Cambridge Univ. Press (2012).

[40] W. Greiner, L. Neise and H. Stöcker, Thermodynamics and statistical mechanics, Springer (2012).

[41] D.J. Griffiths, Introduction to electrodynamics, Cambridge Univ. Press (2017).

[42] D.J. Griffiths and D.F. Schroeter, Introduction to quantum mechanics, Cambridge Univ. Press (2018).

[43] D.J. Griffiths, Introduction to elementary particles, Wiley (2020).

[44] D.J. Griffiths, Revolutions in twentieth-century physics, Cambridge Univ. Press (2012).

[45] J. Harris, Algebraic geometry, Springer (1992).

[46] W.A. Harrison, Solid state theory, Dover (1970).

[47] M.P. Hobson, G.P. Efstathiou and A.N. Lasenby, General relativity, Cambridge Univ. Press (2006).

[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] V.F.R. Jones, Subfactors and knots, AMS (1991).

[53] L.P. Kadanoff, Statistical physics: statics, dynamics and renormalization, World Scientific (2000).

[54] T. Kibble and F.H. Berkshire, Classical mechanics, Imperial College Press (1966).

[55] M. Kumar, Quantum: Einstein, Bohr, and the great debate about the nature of reality, Norton (2009).

[56] T. Lancaster and K.M. Blundell, Quantum field theory for the gifted amateur, Oxford Univ. Press (2014).

[57] L.D. Landau and E.M. Lifshitz, Mechanics, Pergamon Press (1960).

[58] L.D. Landau and E.M. Lifshitz, The classical theory of fields, Addison-Wesley (1951).

[59] L.D. Landau and E.M. Lifshitz, Quantum mechanics: non-relativistic theory, Pergamon Press (1959).

[60] V.B. Berestetskii, E.M. Lifshitz and L.P. Pitaevskii, Quantum electrodynamics, Butterworth-Heinemann (1982).

[61] P. Lax, Linear algebra and its applications, Wiley (2007).
[62] P. Lax, Functional analysis, Wiley (2002).
[63] J.M. Lee, Introduction to topological manifolds, Springer (2011).
[64] J.M. Lee, Introduction to smooth manifolds, Springer (2012).
[65] J.M. Lee, Introduction to Riemannian manifolds, Springer (2019).
[66] M.L. Mehta, Random matrices, Elsevier (2004).
[67] R.K. Pathria and and P.D. Beale, Statistical mechanics, Elsevier (1972).
[68] P. Petersen, Linear algebra, Springer (2012).
[69] P. Petersen, Riemannian geometry, Springer (2006).
[70] W. Rudin, Principles of mathematical analysis, McGraw-Hill (1964).
[71] W. Rudin, Real and complex analysis, McGraw-Hill (1966).
[72] B.M. Peterson and B. Ryden, Foundations of astrophysics, Cambridge Univ. Press (2010).
[73] B. Ryden, Introduction to cosmology, Cambridge Univ. Press (2002).
[74] B. Ryden and R.W. Pogge, Interstellar and intergalactic medium, Cambridge Univ. Press (2021).
[75] D.V. Schroeder, An introduction to thermal physics, Oxford Univ. Press (1999).
[76] B. Schutz, A first course in general relativity, Cambridge Univ. Press (2009).
[77] J. Schwinger, Einstein's legacy: the unity of space and time, Dover (1986).
[78] J. Schwinger, L.L. DeRaad Jr., K.A. Milton and W.Y. Tsai, Classical electrodynamics, CRC Press (1998).
[79] J. Schwinger and B.H. Englert, Quantum mechanics: symbolism of atomic measurements, Springer (2001).
[80] I.R. Shafarevich, Basic algebraic geometry, Springer (1974).
[81] R. Shankar, Fundamentals of physics I: mechanics, relativity, and thermodynamics, Yale Univ. Press (2014).
[82] R. Shankar, Fundamentals of physics II: electromagnetism, optics, and quantum mechanics, Yale Univ. Press (2016).
[83] R. Shankar, Principles of quantum mechanics, Springer (1980).
[84] R. Shankar, Quantum field theory and condensed matter: an introduction, Cambridge Univ. Press (2017).
[85] A.M. Steane, Thermodynamics, Oxford Univ. Press (2016).
[86] A.M. Steane, Relativity made relatively easy I, Oxford Univ. Press (2012).
[87] A.M. Steane, Relativity made relatively easy II, Oxford Univ. Press (2021).
[88] J.R. Taylor, Classical mechanics, Univ. Science Books (2003).
[89] N.H. Temperley and E.H. Lieb, Relations between the "percolation" and "colouring" problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the "percolation" problem, *Proc. Roy. Soc. London* **322** (1971), 251–280.
[90] M. Tsamparlis, Special relativity, Springer (2010).
[91] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press (1955).
[92] S. Weinberg, Foundations of modern physics, Cambridge Univ. Press (2011).

[93] S. Weinberg, Lectures on quantum mechanics, Cambridge Univ. Press (2012).

[94] S. Weinberg, Lectures on astrophysics, Cambridge Univ. Press (2019).

[95] S. Weinberg, Cosmology, Oxford Univ. Press (2008).

[96] H. Weyl, The theory of groups and quantum mechanics, Princeton Univ. Press (1931).

[97] H. Weyl, The classical groups: their invariants and representations, Princeton Univ. Press (1939).

[98] H. Weyl, Space, time, matter, Princeton Univ. Press (1918).

[99] E. Witten, Quantum field theory and the Jones polynomial, *Comm. Math. Phys.* **121** (1989), 351–399.

[100] B. Zwiebach, A first course in string theory, Cambridge Univ. Press (2004).