INTRODUCTION TO RELATIVITY AND QUANTUM MECHANICS

A crash course for Maths students

JOSÉ ANTONIO VALLEJO

Facultad de Ciencias, UASLP

http://galia.fc.uaslp.mx/~jvallejo

June 5, 2018

ABSTRACT

These notes are meant to be the first part of a 3 sessions introduction to Quantum Field Theory for undergraduate students of Mathematics with no previous knowledge of Physics beyond the elementary ideas taught in a high school course. To compensate for this, we assume a certain mathematical proficiency. The contents were presented at the First Summer School on Arithmetics, p−adic Analysis and Mathematical Physics held in Bogotá (Colombia), June, 5–8, 2018. Due to time constraints, the notes have not been revised, and it is very likely they contain some errors (both typographical and conceptual). Use them with care.

CONTENTS

1 Classical (non-quantum) Physics 2
- 1.1 Relativity . . . 2
- 1.2 Simultaneity and Causality . . . 4
- 1.3 Energy and momentum . . . 7
- 1.4 Fields . . . 9
- 1.5 Exercises . . . 12

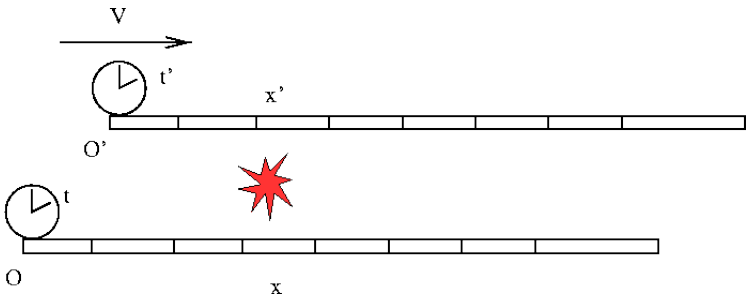
2 Quantum Physics 13
- 2.1 The birth of the quantum . . . 13
- 2.2 Postulates of Quantum Mechanics . . . 16
- 2.3 Wave equations: Schrodinger and Klein-Gordon . . . 19
- 2.4 The harmonic oscillator . . . 22
- 2.5 Exercises . . . 26

1 CLASSICAL (NON-QUANTUM) PHYSICS

1.1 Relativity

We must begin by considering the structure of *spacetime*. In Newtonian physics it is just $M = \mathbb{R} \times \mathbb{R}^3$ as a product vector space endowed with two distances $d_{\mathbb{R}}(t,s) = |t-s|$ and $d_{\mathbb{R}^3}(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$, where $\|\cdot\|$ is the Euclidean norm. This just means that space and time actually are separated entities, not influencing each other.

Now take into account the Einstein postulate about the constancy of light speed in vacuum c (which is indeed a fact experimentally verified). For simplicity, let us restrict our attention to the case of a plane $\mathbb{R} \times \mathbb{R} \subset \mathbb{R} \times \mathbb{R}^3$ with coordinates (t,x). If we set up two laboratories L and L′, arbitrarily considering one of them (L) fixed and the other (L′) displacing with respect to the first with constant velocity v to the right, and assuming that for $t = 0 = t'$ the two laboratories had their respective origins O and O′ located at $x = 0 = x'$, firing up a beam of light from the origin of L it will reach a certain point after a time t has elapsed in that laboratory. In laboratory L′, the same light beam will reach a point characterized by the coordinate value x', after a time t'.



The constancy of c implies that

$$\frac{x}{t} = c = \frac{x'}{t'}$$

or $c^2t^2 - x^2 = 0 = c^2t'^2 - x'^2$. In other words, the quantity defined by $\Delta^2 \doteq c^2(\text{time})^2 - (\text{space})^2$ is an *invariant* for any observer in a laboratory moving with constant velocity (that is, without acceleration) with respect to one fixed as a reference. To characterize the invariant Δ^2, notice its similarity with the square of lengths of vectors in an Euclidean space. The presence of the minus sign can be taken into account if we define an *inner product* in $M = \mathbb{R} \times \mathbb{R}^3$ *with coordinates* (ct, x^1, x^2, x^3)[†] by the symmetric matrix

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}, \tag{1}$$

so the product of $x = (ct, \mathbf{x})$ and $y = (ct', \mathbf{y})$ will be

$$\langle x, y \rangle = c^2tt' - \mathbf{x} \cdot \mathbf{y},$$

† There is no problem in including the factor c in the definition of the first coordinate as, by Einstein's postulate, it is a universal, fixed number.

where $\cdot$ is the Euclidean scalar product in $\mathbb{R}^3$. In particular, the squared 'length' of a vector $x \in M$ is precisely Δ^2:

$$x^2 = x^\top \eta x = c^2 t^2 - \mathbf{x}^2 = \Delta^2 \,. \tag{2}$$

The pair $(\mathcal{M}, \eta)$ is called *Minkowski spacetime*, or just spacetime for short. Then, the fact that $\Delta^2 = c^2 t^2 - x^2$ is the same for any observer in a laboratory satisfying the conditions stated above, can be rephrased by saying that the admissible laboratories (let us call them *reference frames* as is customary in Physics) are defined by *linear* transformations $T : M \to M$ that preserve the length squared of vectors in (M, η). Thus, admissible reference frames are given by linear transformations such that $T^\top \eta T = \eta$. It is an easy exercise in linear algebra to determine the matrices $[T]$; in fact, if we write the two dimensional case explicitly as

$$[T] = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} ,$$

we readily find the conditions

$$\begin{aligned} T_{11}^2 - T_{12}^2 &= 1 \\ T_{11} T_{12} - T_{21} T_{22} &= 0 \\ T_{12}^2 - T_{22}^2 &= -1 \,. \end{aligned}$$

Recaling the definition and properties of the hyperbolic sine and cosine functions, we have the explicit realization

$$[T] = \begin{pmatrix} \cosh \xi & \sinh \xi \\ \sinh \xi & \cosh \xi \end{pmatrix} ,$$

for some $\xi \in \mathbb{R}$ that parametrizes the transformation (much in the same way as the angle θ parametrizes rotations in Euclidean space).

The transformation between coordinates in the reference frames (laboratories) L and L′ is thus

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh \xi & \sinh \xi \\ \sinh \xi & \cosh \xi \end{pmatrix} \cdot \begin{pmatrix} ct \\ x \end{pmatrix} \tag{3}$$

Now, if L′ is moving to the right of L with velocity v, the point $x' = 0$ in L′ will correspond to the point $x = vt$ in L. This suffices to determine the relation between ξ and v using (3):

$$ct \sinh \xi + vt \cosh \xi = 0 \,,$$

hence $\tanh \xi = -v/c$. It is customary in Physics to denote

$$\beta \doteq \frac{v}{c} \quad \text{and} \quad \gamma \doteq \frac{1}{\sqrt{1-\beta^2}} \,, \tag{4}$$

so, taking into account the hyperbolic trigonometric identities relating cosh, sinh and tanh (all of them derived from $\cosh \xi - \sinh \xi = 1$) we get

$$\sinh \xi = -\beta\gamma \quad \text{and} \quad \cosh \xi = \gamma \,.$$

Putting all of this together, we get the following explicit expressions for the so-called *Lorentz transformations* in the 2−dimensional case:

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \gamma \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix} \cdot \begin{pmatrix} ct \\ x \end{pmatrix} \tag{5}$$

Notice that, as a consequence of (4) and another postulate of Einstein (that no physical motion can proceed at a speed faster than c), the range for β is actually $[-1, 1]$, and that of γ, $[1, \infty[$.

Remark 1. It is usual, to lighten the notation, to take units in which $c = 1$ and we will do so in what follows. Those are called *natural units*.

1.2 Simultaneity and Causality

A straightforward consequence of the geometric structure of spacetime $\mathcal{M}$, is the fact that simultaneity is relative, that is, it is a notion depending on the reference frame chosen. For a given reference L, two points (or, in the physicists' language, *events*) $x = (t, \mathbf{x}), y = (s, \mathbf{y}) \in \mathcal{M}$ are said to be *simultaneous* if $s = t$ (as measured by L).

In the 2−dimensional case, if L′ is another reference moving with respect to L with relative velocity v, the Lorentz transformations (5) give the L′ coordinates for two events $A = (t, x)$ and $B = (t, y)$ simultaneous in L:

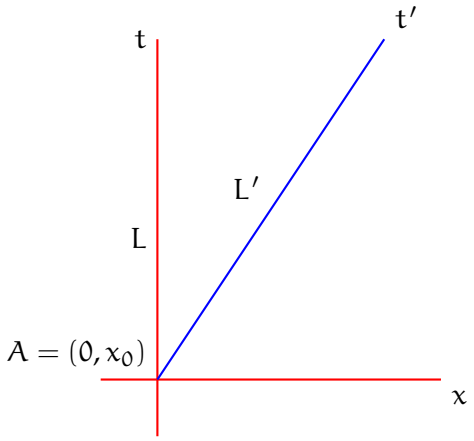
$$A' = (\gamma t - \beta\gamma x, -\beta\gamma t + \gamma x) \text{ and } B' = (\gamma t - \beta\gamma y, -\beta\gamma t + \gamma y)\,.$$

Notice that, unless $x = y$ (that is, the events A and B are actually the same), then A′ and B′ are no longer simultaneous. Their difference in the time coordinate will be

$$|\gamma t - \beta\gamma x - \gamma t + \beta\gamma y| = \beta\gamma |y - x| \neq 0\,.$$

Of course, this consequence is equally valid in the full 4−dimensional case.

For a given reference frame L, the events occuring simultaneously with a fixed one $(t_0, \mathbf{x}_0)$ are given by the hyperplane $t = t_0$. But those events are different for an observer attached to a reference L′ in relative motion with respect to L. To understand what happens, let us restrict ourselves again to the 2−case. The motion of L′ with respect to L is given by a straight line with slope $v = x/t$. In the following figure we represent L (in red), the line of simultaneity at $t = 0$ (in red), and L′ (in blue), all from the point of view of L (so its position, at the origin, coincides with its own time axis):



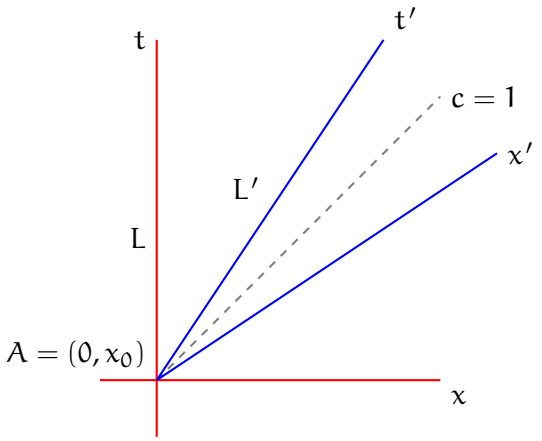
Notice that we are taking $c = 1$ and the relative speed v cannot be greater than c, so the angle of the line representing L' must be greater than $\pi/4$. What will be the lines of simultaneity of the events *in the reference* L'? For the L' observer, she is at rest with respect to herself, so her lines of simultaneity are just 'horizontal' lines. So, the question is how those simultaneity lines are seen by L. For instance, the simultaneity line of A in L' is just the set of points with $t' = 0$. By computing the inverse transformation of (5) we can find that the coordinates (t, x) in L of these points satisfy

$$t = \gamma\beta x'$$
$$x = \gamma x',$$

so they are represented by the line

$$t = \beta x.$$

Recalling that $\beta = v/c$ and that we are taking $c = 1$, we see that $t = \beta x = vx$ is the reflection of the line $t = x/v$ along the bisectrix. Thus, the diagram giving the lines of simultaneity for L and L' is



Simultaneity enters into the idea of *causality*. Consider an observer that detects two different but simultaneous events; they occur at the same time but at different places. Under these circumstances there is no way they can interfere (in legal terms any one of them could say it has an alibi for whatever occurred at the other), so the observer must conclude that there is no cause-effect relation between them. In physical terms, we say that one event has a *causal relation* to another if there exists a reference frame in which it is possible to transmit a light signal from one of them to the other. If $A = (t, \mathbf{x}), B = (t', \mathbf{y})$ are the respective coordinates, this definition implies that the relation

$$\|\mathbf{x} - \mathbf{y}\|^2 \leqslant |t - t'|^2$$

(expressing the fact that the required velocity for travelling from A to B is less than $c = 1$) must hold. We can restate this in terms of the quantity Δ corresponding to the vector $A - B$, which receives the name of *spacetime interval* between A and B:

$$\Delta^2 = |t - t'|^2 - \|\mathbf{x} - \mathbf{y}\|^2 \geqslant 0,$$

that is: two events are causally related if their spacetime interval is positive. We also say that the events are *timelike* separated. Analogously, A and B are *spacelike* separated if $\Delta^2 = |t - t'|^2 - \|\mathbf{x} - \mathbf{y}\|^2 < 0$.

As it has been stated, the definition of causality seems to depend on the reference frame. However, the following result holds.

Theorem 1. *If two events are timelike separated in a reference frame* L*, then they are timelike separated in any other* L′ *moving at a constant speed* v *with respect to* L.

Proof. Let $A = (t, \mathbf{x}), B = (s, \mathbf{y})$ be the coordinate of the events in L, and $A = (t', \mathbf{x}'), B = (s', \mathbf{y}')$ the corresponding coordinates in L′. We know that the difference vectors $u = (t - s, \mathbf{x} - \mathbf{y})$ and $u' = (t' - s', \mathbf{x}' - \mathbf{y}')$ are related through $u' = \Lambda u$, where Λ is the matrix appearing in (5):

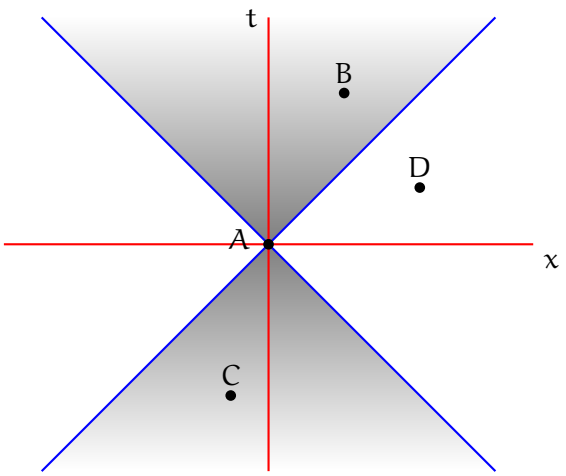
$$\Lambda = \gamma \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix} .$$

But then we can compute the spacetime interval corresponding to v' in L′:

$$\Delta'^2 = u'^{\top} \eta u' = u^{\top} \Lambda^{\top} \eta \Lambda u = u^{\top} \eta u = \Delta^2 ,$$

so if $\Delta^2 \geqslant 0$, then $\Delta'^2 \geqslant 0$. □

Spacetime diagrams provide a nice geometrical way of representing causality. Take on point A and trace two bisectrices of the plane crossing at it, with slope 1. These represent light rays emanating from A. All the points causally connected with A lie inside the interior of the 'cones' so defined: the points in the *forward or future cone* represent those events that can be caused by A, while those in the *backward or past cone* represent those that can be a cause for A.



Thus, in the preceding diagram, A is causally connected to B and C, but not to D.

The relation between simultaneity and causality is further clarified by the following result.

Proposition 1. *Let* $A = (t, \mathbf{x}), B - (s, \mathbf{y})$ *be two spacelike separated events in a reference frame* L*. Then, it is possible to find a transformation* Λ *from* L *to a new reference frame* L′ *such that the events are simultaneous.*

Proof. For simplicity, we will work in the 2−dimensional case. The proposed transformation law is, according to (5),

$$\begin{pmatrix} t'-s' \\ x'-y' \end{pmatrix} = \gamma \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix} \begin{pmatrix} t-s \\ x-y \end{pmatrix}.$$

From the equality of the temporal coordinates we deduce that the sought Λ must satisfy

$$t'-s' = \gamma(t-s) - \gamma\beta(x-y),$$

that is, it is determined by the parameter $\beta = v$ (remember, $c = 1$) such that

$$\beta = \frac{t-s}{x-y}.$$

For this to make sense, it must happen that $\beta < 1$. But this is precisely guaranteed by the fact that A and B are spacelike separated in L, for this implies that $(t-s)^2 - (x-y)^2 < 0$, or $|(t-s)/(x-y)| < 1$. □

Remark 2. Notice, incidentally that trying to make simultaneous two timelike separated events using the technique of the preceding proposition, would require a parameter $\beta > 1$ and thus is not physically realizable.

1.3 Energy and momentum

In this section we consider a parametrized curve in spacetime, $\alpha : I \subset \mathbb{R} \to \mathcal{M}$, describing the motion of a particle with mass m. The parameter of the curve, of course, can be chosen arbitrarily, but there is a natural choice given by the arc-length with respect to the quadratic form η. In physics, instead of 'arc-length with respect to the quadratic form η' we speak about *proper time* (we will see the reason for this in a moment). Thus, the parameter τ is called the *proper time* of the particle if

$$\sqrt{\left(\frac{d\alpha}{d\tau}\right)^{\top} \eta \, \frac{dffc}{d\tau}} = 1, \tag{6}$$

for all $\tau \in I$.

Remark 3. Notice that this definition makes sense, as the vector $d\alpha/d\tau$ is always timelike for a physical trajectory, and thus the square root gives a real number.

The measures done in a laboratory L are usually expressed in terms of a parameter t chosen by the observer, and this choice is made *before* considering any curve, so we can not expect that once a motion is observed, t coincides with its proper time τ. Hence, the problem of comparing the description of a trajectory using both parameters arise. This problem is related to the kinematic magnitudes of energy and momentum through the following definition: let $\alpha : I \subset \mathbb{R} \to \mathcal{M}$ be a curve parametrized by its proper time. Then, its *energy-momentum* 4−vector is 'mass times velocity', that is,

$$p = m\frac{d\alpha}{d\tau}.$$

In particular, the *energy* of the particle is defined as the first component[†] of p, and the *relativistic 3—momentum* (or, simply, the relativistic momentum) is the 'spatial' part of p, denoted $\mathbf{p}$. Thus, $p = (E, \mathbf{p})$.

Let us consider the point of view of the observer in L, who uses the parameter t to describe the curve c. What she can measure is

$$m\frac{d\alpha(t)}{dt},$$

a vector that can be related to the energy-momentum by applying the chain rule and noticing that the mass m is an invariant:

$$m\frac{d\alpha(\tau)}{d\tau} = m\frac{d\alpha(t)}{dt}\frac{dt}{d\tau}.$$

For a particle moving with constant speed v in L, the functional dependence between τ and t is given by (5), which in this case reads $t = \gamma\tau + \beta x$, hence in this case

$$m\frac{d\alpha(\tau)}{d\tau} = m\frac{d\alpha(t)}{dt}\gamma,$$

or

$$p = m\gamma\frac{d\alpha(t)}{dt}.$$

Notice that a particle moving with constant speed v in L can be parametrized by $\alpha(t) = (t, \mathbf{v})$. Replacing this in the preceding equation gives at once

$$E = m\gamma \tag{7}$$

$$\mathbf{p} = m\gamma\mathbf{v}. \tag{8}$$

The first equation (7) is usually written including the c factors (standard metric units). In this case, we would write, according to the definition, $E/c = m\gamma c$, or

$$E = m\gamma c^2.$$

This is the total energy for a particle of mass m moving with speed v. It can be further developed by taking the limit in which $\beta = v/c \to 0$ (that is, the limit $v \leqq c$), for then a Taylor expansion of γ gives

$$E = mc^2\left(1 + \frac{\beta^2}{2} + O(\beta^3)\right) \simeq mc^2 + \frac{1}{2}mv^2.$$

This famous formula by Einstein, shows that even with $v = 0$, the particle has a *rest energy* equal to $E_0 = mc^2$. On the other hand, (8) gives the expression of the relativistic momentum as a modification of the non-relativistic one $m\mathbf{v}$:

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

To finish this brief excursion into the relativistic energy, we note that the square norm of the energy-momentum $p = (E, \mathbf{p})$ can be computed in two ways. First, of course, as $p^2 = E^2 - \|\mathbf{p}\|^2$, but also as

$$p^2 = \left(m\frac{dc}{d\tau}\right)^2 = m^2,$$

† We are assuming natural units here. If standard metric units are used, then the energy is the first component multiplied by c.

where use has been made of the definition of proper time τ (6). Thus

$$p^2 = E^2 - \|\mathbf{p}\|^2 = m^2 \,. \tag{9}$$

If we work with standard metric units, we need to include the corresponding c factors in (9), which will then read $p^2 = E^2/c^2 - \|\mathbf{p}\|^2 = m^2c^2$, hence the also well-known formula

$$E^2 = m^2c^4 + \|\mathbf{p}\|^2c^2 \,.$$

1.4 Fields

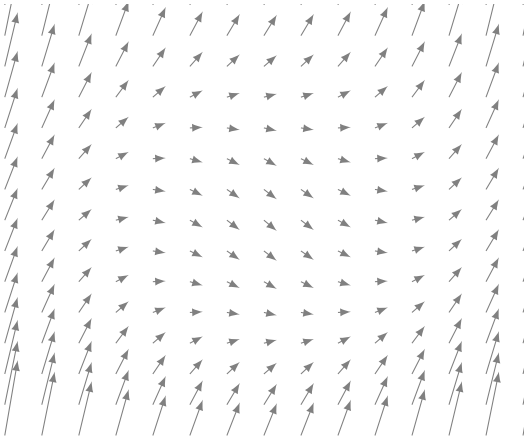
Newton's equations are well-suited to study the motion of pointlike particles (and systems of them), but in advanced Physics it is far more common to deal with *fields*. For instance, the electromagnetic field (being generated by cell phones) and the gravitational are all around us, and they need other tools to be properly studied, which are briefly revised in this section.

Particles can be located in space, and their motion in spacetime $\mathcal{M}$ (which we will parametrize as $\mathcal{M} = \mathbb{R} \times \mathbb{R}^3$ with coordinates $(t, \mathbf{x})$ for time[†] and space) is described by curves $\alpha \in \mathcal{C}(I; \mathbb{R}^3)$, with $I \subset \mathbb{R}$ an open interval. Thus, the motion of a Newtonian particle is obtained by solving the equation (with appropriate initial conditions)

$$\frac{d}{dt}p = F(\alpha(t), \dot{\alpha}(t)) \,,$$

where p is the relativistic energy-momentum, $F : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ is the *relativistic force*, and the dot in $\dot{\alpha}$ denotes the derivative with respect to t. On the other hand, fields are not localized, they spread all over the space, and can not be described by curves. The electromagnetic field, for instance, is given by the data of two vectors $\mathbf{E}(t, \mathbf{x})$, $\mathbf{B}(t, \mathbf{x})$ simultaneously at each point of space, although these two vectors vary with time as the notation suggests. Thus, instead of a curve we would need something similar to a mapping $F : \mathcal{M} \to \mathcal{M} \times \mathcal{M}$.

Sometimes physicists say that giving a physical field is just putting an arrow at each point of space. Thus the velocity field of a fluid is just the assignment of the velocity vector of the fluid at each position it occupies, as graphically illustrated in the figure below.



† In this section we will consider a fixed reference frame, t will denote its measured time, and τ will be the proper time along a given trajectory.

In general, a physical field can assign any magnitude to points on space: a *scalar field* assigns just a number, as in the case of a temperature field. Also, we will consider fields defined not only on space, but in the spacetime $\mathcal{M} = \mathbb{R} \times \mathbb{R}^3$.

To describe the evolution of physical entities (both particles and fields), it is useful to introduce the following concepts. Let $U \subset \mathcal{M}$ be an open set. If $\mathbf{x} \in U$, the *tangent space* at $\mathbf{x}$ is the set

$$T_{\mathbf{x}}U \doteq \{\mathbf{x}\} \times \mathcal{M} = \{(\mathbf{x}, \mathbf{v}) : \mathbf{v} \in \mathcal{M}\}.$$

The elements $(\mathbf{x}, \mathbf{v})$ are the *tangent vectors* at $\mathbf{x}$. When there is no risk of confusion about the point we are considering, we will simply speak about the tangent vector $\mathbf{v}$. One can safely think of $T_{\mathbf{x}}U$ as a copy of $\mathcal{M}$ with its origin at $\mathbf{x}$. The *tangent bundle* TU is the disjoint union of all the tangent spaces:

$$TU \doteq \bigcup_{\mathbf{x} \in U} T_{\mathbf{x}}\mathcal{M}.$$

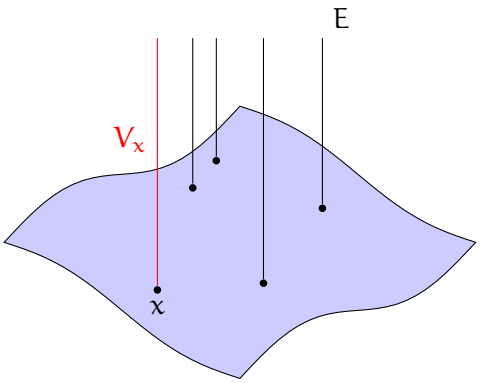
The tangent bundle comes equipped with a canonical *projection*, $\pi : TU \to U$, such that $\pi(\mathbf{x}, \mathbf{v}) = \mathbf{x}$. A *local section* of the tangent bundle is a mapping $\phi : U \to TU$ such that it assigns to each point $\mathbf{x}$ a tangent vector at $\mathbf{x}$, in other words,

$$\pi \circ \phi = \mathrm{Id}.$$

The preceding construction can be generalized by replacing in the definition of tangent space at a point the vector space $\mathcal{M}$ by any other space V (real or complex). A *vector bundle with fiber* V over $\mathcal{M}$ is then a quadruple $(E, \pi, \mathcal{M}, V)$ such that $p : E \to$ is smooth, for each $x \in \mathcal{M}$ there exists an open set $x \in U \subset \mathcal{M}$ satisfying the condition

$$\pi^{-1}(U) \simeq U \times V,$$

and the so-called *fiber* over the point $x \in \mathcal{M}$, the set $\pi^{-1}(\{x\})$ is a vector space isomorphic to V, that is, $\pi^{-1}(\{x\}) \doteq V_x \simeq V$. Another way to say that is by requiring that the isomorphism $\varphi : p^{-1}(U) \to U \times V$ satisfies $\mathrm{pr}_1 \circ \varphi = \pi|_{\pi^{-1}(U)}$. Thus, the vectors in the fiber V_x really are 'attached' to x in a precise way.



Remark 4. at This definition may seem overly complicated, and indeed it is. However, the bundles we will use are *trivial* in the following sense: the *total space* E will have the form $E = \mathcal{M} \times V$ for an appropriate V, and the projection

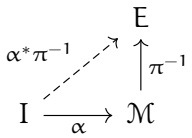
π will be the canonical one $\pi = \mathrm{pr}_1$. There is no loss of generality in doing so, because $\mathcal{M}$ is contractible (and it is known from algebraic topology every vector bundle over a contractible space is trivial).

A *local section* of a vector bundle $(E, \pi, \mathcal{M}, V)$ is a mapping $\sigma : U \subset \mathcal{M} \to E$ such that $\pi \circ \sigma = \mathrm{Id}_U$, where $U \subset \mathcal{M}$ is an open set. The section is a *global* one if $U = \mathcal{M}$. A physical *field* is just a section of an appropriate bundle over spacetime $\mathcal{M}$.

Remark 5. A general characteristic of sections on trivial bundles is that they always have the form $\sigma = (\mathrm{Id}, \widetilde{\sigma})$ for some $\widetilde{\sigma} : U \to V$. It is then customary to call $\widetilde{\sigma}$ 'the field', omitting the trivial identity factor.

Example 1. A complex scalar field can be described as follows. Take $E = \mathcal{M} \times \mathbb{C}$ with $\pi = \mathrm{pr}_1$. A local section of this bundle is a mapping $\sigma : U \to U \times \mathbb{C}$ having the form $\sigma(x) = (x, \phi(x))$, where for each $x \in U \subset \mathcal{M}$, $\phi(x) \in \mathcal{C}$ is a complex number. The field itself can be then identified (following Remark 5) with a mapping $\phi : U \to \mathbb{C}$. △

Example 2. The bundle formalism is quite flexible and, with some modifications, allows the treatment of particle dynamics. To this end, we introduce the notion of *pullback of a vector bundle along a curve*. Given a vector bundle $(E, \pi, \mathcal{M}, V)$ and a curve $\alpha : I \subset \mathbb{R} \to \mathcal{M}$, we construct a new vector bundle *over* I, $(\alpha^* E, \alpha^* \pi, I, V)$, by defining its fibers as $\alpha^* \pi^{-1}(t) \doteq \pi^{-1}(\alpha(t))$, that is, the fiber over $t \in I$ is just the fiber over its image $\alpha(t)$ in the original bundle on $\mathcal{M}$. Graphically:



Consider a particle moving in the spacetime $\mathcal{M}$, following a parametrized curve $\alpha(t)$ (in some reference frame L). Its velocity can be considered as a section of the pullback of the tangent bundle $T\mathcal{M}$ along α, acting as

$$\sigma(t) = (\alpha(t), \dot{\alpha}(t)) .$$

Sometimes this is also called the *canonical lifting* of α to a curve on $T\mathcal{M}$. △

Example 3. The electromagnetic field is described by a pair of vectors $\mathbf{E}(t, \mathbf{x})$, $\mathbf{B}(t, \mathbf{x})$, varying in the spacetime coordinates $(t, \mathbf{x})$. To conform with the literature, instead of $(t, \mathbf{x})$ we will employ a unified notation putting $x^0 = t$ and $x^\mu = (t, \mathbf{x}) = (x^0, x^i)$, with ranges $0 \leqslant \mu \leqslant 3$ and $1 \leqslant i \leqslant 3$.

We want to construct a bundle whose sections represent the electromagnetic field. Thus, we should place a pair of vectors at each point of spacetime, and our first guess would be to consider the trivial bundle with total space $E = \mathcal{M} \times \mathbb{R}^6$. There is nothing wrong with that, but it is convenient to use the isomorphism $\mathbb{R}^6 \simeq \Lambda^2(\mathcal{M})^*$ (the space of skew-symmetric bilinear forms on spacetime) and describe the electromagnetic field locally by the skew-symmetric matrix

$$[F_{\mu\nu}] = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} \tag{10}$$

Notice that $F_{\mu\nu}$ actually has a dependence on points $x^\mu \in \mathcal{M}$, that

$$F_{0i}(x^\mu) = E_i(x^\mu)\,,$$

and, for $i < j$,

$$F_{ij} = \varepsilon_{ijk} B_k\,,$$

where ε_{ijk} is the totally antisymmetric Levi-Cività symbol. When viewed as an alternating second-order covariant tensor (that is, a 2–form) on $\mathcal{M}$, (10) is called the *Faraday form*.

Putting all together, the electromagnetic fields can be defined as sections of the trivial vector bundle with total space $E = \mathcal{M} \times \Lambda^2(\mathcal{M})^*$, so we identify them (by Remark 5) with mappings $F : \mathcal{M} \to \Lambda^2(\mathcal{M})^*$. Such mappings are called *differential 2–forms* on $\mathcal{M}$, and the space of all of them is denoted $\Omega^2(\mathcal{M})$. △

1.5 Exercises

In all the exercises, $\mathcal{M} = \mathbb{R} \times \mathcal{R}^3$ is the spacetime, endowed with the inner product (1).

1. A vector $x \in \mathcal{M}$ is called *timelike* if $\langle x \rangle > 0$, *spacelike* if $\langle x \rangle < 0$, and *null* if $\langle x \rangle = 0$. Prove that if x is timelike, any y orthogonal to it, $y \in \{x\}^\perp$, is spacelike, and that the following decomposition holds:

$$\mathcal{M} = \mathbb{R}x \oplus \{x\}^\perp\,.$$

2. Let $x, y \in \mathcal{M}$ be two timelike vectors, that is, $\langle x, x \rangle\,, \langle y, y \rangle > 0$. Prove that

$$|\,\langle x, y \rangle\,| \geqslant |x||y|\,,$$

where, on the right hand side, $|x| = \sqrt{\langle x, x \rangle}$ (and similarly for y).

3. Let $x, y \in \mathcal{M}$ be two timelike vectors. Prove that

$$|x + y| \geqslant |x| + |y|\,.$$

4. Here is the statement of the 'twin paradox' in the Encyclopedia Britannica[†]:

> Suppose that one of two identical twin sisters flies off into space at nearly the speed of light. According to relativity, time runs more slowly on her spacecraft than it does on Earth; therefore, when she returns to Earth, she will be younger than her Earth-bound sister. But in relativity, what one observer sees as happening to a second one, the second one sees as happening to the first one. To the space-going sister, time moves more slowly on Earth than it does in her spacecraft; when she returns, her Earth-bound sister is the one who is younger. How can the space-going twin be both younger and older than her Earth-bound sister?

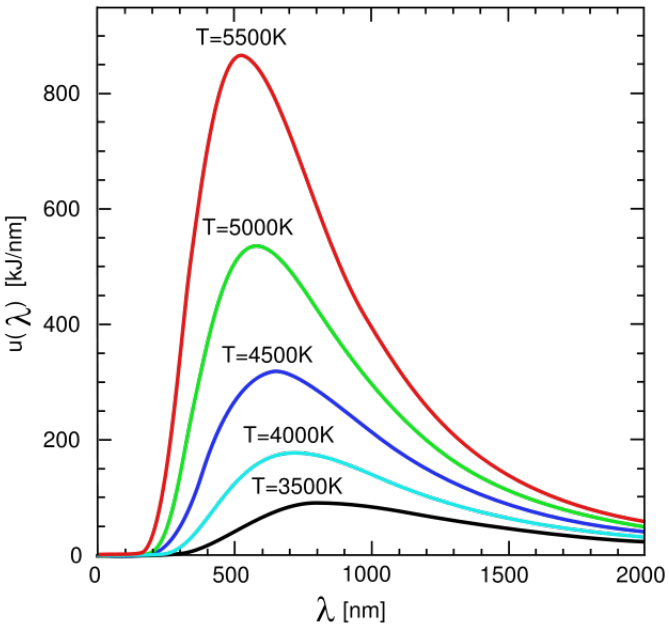
Solve the twin paradox using the preceding exercices.

† Available on line at the URL https://www.britannica.com/science/twin-paradox.

2 QUANTUM PHYSICS

2.1 The birth of the quantum

By the end of the 19th century, physicists were studying the electromagnetic radiation emitted by a heated body. They assumed that, in the process of heating, the energy of the body increased (being absorbed by its atoms from the heat source) until a state of thermal equilibrium was reached. From that moment on, all the absorbed energy was supposed to be completely radiated in the form of light. That light is composed of waves each one of which had a characteristic wavelength, which is related to its color in the case of visible light. The main interest was to explain the observed distribution of the energy radiated at the equilibrium given by temperature T, in terms of the wavelengths. That is: once equilibrium was reached, at a certain temperature T, light was radiated in different wavelengths, but not the same energy was dedicated to each one of them: very short and very large wavelengths were 'less radiated'. The resulting graph with the distribution of the density of radiated energy (per volume unit) was the same independently of the composition of the body, and it was called the *black body spectrum*[†]. It looks like that:



Classical Physics modeled the process of energy absorption by supposing that the molecules of the black body behaved like harmonic oscillators. This hypothesis was based on the fact that the entire energy exchange occurred

† The adjective 'black' comes from one of the experimental settings to study radiation: a hollow metallic cylinder with its inner walls painted on black, with a small hole perforated in its surface. This body absorbs all the energy communicated to it by heating, and the quantity measured is the radiation emanating from the hole. The setting is very similar to a room with a small window in a tall building. To see its interior, light coming from the sun and reflected on the furniture (say) has to escape through the window and reach you on the ground, but it is very improvable that this happens, due to the small area of the window compared to the room walls. Thus, most of the incoming radiation is absorbed by the room, nothing is seen and the window looks black, hence the name.

at a thermal equilibrium: is we have a classical system governed by some Newtonian equations of type

$$\mathbf{F} = -\text{grad}V\,,$$

for some autonomous (that is, time-independent) but otherwise arbitrary potential $V = V(\mathbf{x})$, to be at equilibrium means that the total force is zero, so there is a configuration $\mathbf{x}_0$ for which

$$\mathbf{F}(\mathbf{x}_0) = \mathbf{0} = -\text{grad}V(\mathbf{x}_0)\,.$$

To better grasp the implications, let us work in one dimension, with $\mathbf{x}_0 = x_0$. Then, the preceding equation states that x_0 is a critical point for the potential energy V. Making first a translation through the change of coordinates $\xi = x - x_0$ and expanding by Taylor up to second order, we get

$$V(\xi) = V(0) + \frac{V''(0)}{2}\xi^2\,.$$

A further rescaling making $V(0) = 0$ finally leads to

$$V(\xi) = \frac{V''(0)}{2}\xi^2\,,$$

which is the potential for a harmonic oscillator of natural frequency

$$w = \sqrt{\frac{V''(0)}{m}}\,.$$

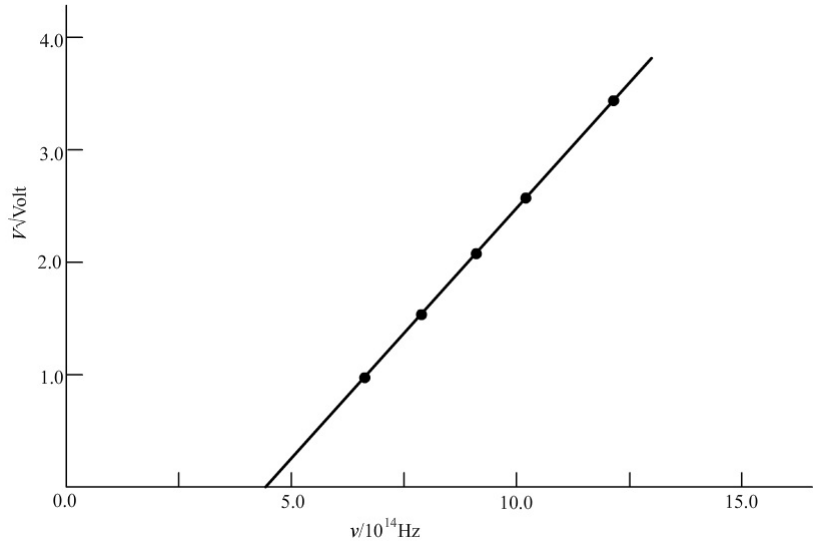
Remark 6. Notice that the above reasoning is completely general, valid for any potential: near an equilibrium, up to second order *any* system can be approximated by a harmonic oscillator.

A classical harmonic oscillator can absorb and release energy in a continuous way (think of the extension and compression of a spring), but Planck realized, in the winter of 1900, that in order to deduce the correct expression for the radiated energy density $u(\lambda)$ it is necessary to accept that the exchange of energy between the radiation and the molecules-oscillators of the black body only takes place for certain values of E, those multiples of a certain fundamental energy $E_0 = h\nu$. Here ν is a frequency (related to the wavelength of light by $\nu\lambda = c$), and h is a constant of Nature whose value is $h = 6.62607004 \cdot 10^{-34}\,m^2\,kg/s$. It is called the *Planck constant*.

Remark 7. It is very important to emphasize the extremely small value of h. Although the energy exchange proceeds jumping from $h\nu$ to $2h\nu$ and so on, the gaps between these values is so small that is not perceived in the macroscopic world. But the fact that the process is essentially a discrete one has very profound implications in Physics.

The idea that energy was exchanged in small parcels, or *quanta* (singular, *quantum*), was revolutionary but not well received. It clashed with the well-established knowledge gained from the classical principles based on the continuous character of Nature and its processes, and was forgotten for four years, until Einstein resurrected the idea to explain the photoelectric effect.

Experimentally, it is found that when a beam monochromatic light (of frequency ν) is directed towards the surface of a metallic substance, electrons are emitted from the surface. In itself, this effect is not strange at all: the electrons on the surface absorb the energy of the incident light, start to vibrate with increasing amplitudes and eventually escape from the attractive potential of the metallic nuclei around them. The problem is that, following the ideas of classical Physics, the energy carried by the light is proportional to the intensity of the beam, not to its frequency. More incident light extracts more electrons, and this should happen for each frequency. Experimental data, however, shows the following:



It is clearly seen that there exists a threshold frequency (about $4.5 \cdot 10^{14}$Hz), below it, no matter the intensity of light falling on the metal, there is no emission of electrons. With incident frequencies above the threshold, the energy of the emitted electrons is proportional to the frequency and independent of the intensity.

Einstein's explanation is that $W = h\nu_0$, where ν_0 is the threshold frequency, is the work required to extract any electron from the attracting potential created by the surrounding, positively charged, metallic nuclei. Light of frequency ν is a beam of *photons*, each one carrying an amount of energy $E = h\nu$. When one of these photons collide with a surface electron, transfers its energy to it. Thus, if $\nu \leqslant \nu_0$ this energy is not enough for the electron to cross the potential barrier and there is no emission. On the other hand, if $\nu > \nu_0$, a fraction $h\nu_0$ of the energy is invested in crossing the potential barrier, and the remaining $K = h(\nu - \nu_0)$ becomes the kinetic energy of the emitted electron. THe electronic current so produced is then linearly dependent on the frequency ν.

The phenomena just described made clear that electromagnetic radiation, when interacting with matter, behaves as a beam of particles with individual energy $E = h\nu$, where ν is the radiation frequency. In 1924, De Broglie postulated that this dual character is extended to all entities in Nature: in the same way that radiation of frequency ν can be seen as particles of energy

$E = h\nu$, any particle with linear momentum p (that is, energy $E = p^2/2m$) has associated a wave with wavelength

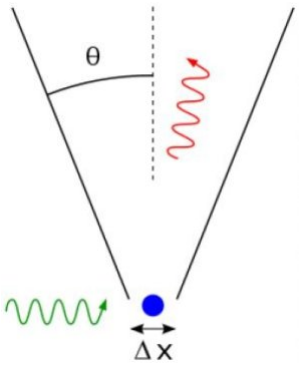
$$\lambda = \frac{h}{p}. \tag{11}$$

This conjecture (called the *particle-wave duality*) was proved true in 1927 by Davisson and Germer, studying the diffraction of slow electrons by a plate of nickel. Einstein received the Nobel price in Physics in 1921, for his explanation of the photoelectric effect, De Broglie was awarded with the Nobel in 1929, and Davisson in 1937.

2.2 Postulates of Quantum Mechanics

Whenever we make a measurement on a physical system, we perturb it. Classical physics assume that these perturbations are negligible and do not affect the outcome (as in the case of measuring the length of a table with a ruler, in which we discard the interaction betweens atoms of the ruler and the wood of the table). This is justified by the fact that measurements in classical physics are always macroscopic, implying values several orders of magnitude greater that the values of the perturbations introduced by interactions at atomic scales. But this is the same reason why those perturbations are relevant when studying a microscopic system.

Example 4. Let us consider the following thought experiment, known as *Heisenberg's microscope*. We have an electron moving on a straight line and we want to determine its position x with a microscope that uses light of wavelength λ and has an angular aperture θ, as in the figure below.



From Optics, we know that the microscope can resolve the position only within an accuracy of

$$\Delta x = \frac{\lambda}{\sin\frac{\theta}{2}} \simeq \frac{2\lambda}{\theta}. \tag{12}$$

But if we observe the electron is because a light photon collided with it, and was reflected towards the microscope lens. This reflection implies that a change in the photon momentum occurred (otherwise it would have continued its rectilinear trajectory to the right, sketched in green in the figure above). By the conservation law, a change in electron momentum, of the

same magnitude, must also have occurred. Applying De Broglie relation (11), the initial total momentum (assuming the electron at rest) is

$$\frac{h}{p} + 0\,,$$

and after the collision, in order for the scattered light to enter the microscope the reflection angle must be in $[-\theta/2, \theta/2]$. The photon momentum, thus, must be comprised after collision in the interval

$$\left[\frac{h}{\lambda} - \frac{h}{\lambda}\sin\frac{\theta}{2}, \frac{h}{\lambda} + \frac{h}{\lambda}\sin\frac{\theta}{2}\right].$$

This has the direct consequence that the transferred momentum can be determined only within a precision of order

$$\Delta p = 2\frac{h}{\lambda}\sin\frac{\theta}{2} \simeq \frac{h\theta}{\lambda}\,, \tag{13}$$

From (12) and (13) we get, finally, the famous *Heisenberg's uncertainty relation*:

$$\Delta x \Delta p \simeq 2h\,.$$

In words: if we increase the resolution of the microscope (making Δx smaller), we increase the imprecision in the measurement of the momentum, so we can no longer guarantee that the electron (that has now acquired a velocity) is standing at rest exactly at the point light was scattered. △

It is very important to notice that the conclusions drawn from Heisenberg's example do not depend on technological details related to the construction of the microscope, or the light used. They refer to limitations inherent to the measurement process which lie at the basis of Quantum Mechanics: when we make a measurement, the system under observation is perturbed in an unpredictable manner and so, *our predictions will have a probabilistic character*.

Remark 8. We have the following two considerations, derived from the experimental facts:

1. To each observation of a physical magnitude, there corresponds a set of real numbers representing the possible outcomes. This set is called the *physical spectrum* of the corresponding magnitude (position, energy, momentum, etc.)

2. If $\mathcal{A}, \mathcal{B}$ represent measurements on a physical system, and $\mathcal{AB}$ is the result of first measuring $\mathcal{A}$ and then $\mathcal{B}$, then, in general,

   $$\mathcal{AB} \neq \mathcal{BA}\,.$$

   Moreover, as the example of Heisenberg's microscope shows, we expect that $\mathcal{AB} - \mathcal{BA}$ be related to the perturbations done by the measurement process on the system.

Von Neumann realized that these properties are similar to those possessed by self-adjoint operators acting on a Hilbert space. He went on to formalize Quantum Mechanics based on the hypothesis that each measurement $\mathcal{A}$ has an associated self-adjoint operator A on a *separable* Hilbert space $\mathcal{H}$. Those operators are called, for obvious reasons, the *observables* of the system. The elements of $\mathcal{H}$ represent the states of the physical system and, due to the particle-wave duality, are called *wave functions* of the system. Notice that, without loss of generality, if the dimension of the space of states is infinite, we can take $\mathcal{H} = L^2(\mathbb{R}^3)$.

Moreover, Von Neumann proposed a set of postulates relating the purely mathematical theory of operators on Hilbert spaces with the physics of microscopic systems.

Postulate I The only possible outcomes for the measurement of $\mathcal{A}$, to which it correspond the operator A, are the eigenvalues of A, that will be denoted a or a_n (when they are discrete).

Postulate II A measurement of $\mathcal{A}$ done upon a system described by a wave function ϕ which is an eigenfunctionof A with corresponding eigenvalue a, so

$$A\phi = a\phi\,,$$

will give as result a *always*, under any circumstance (that is, with probability 1).

Postulate III If we have a set of *identical* states, each one of them described by a wave function $\phi \in \mathcal{H}$, and we measure the observable A on each one of them, the average value we obtain is given by

$$\overline{a}_\phi = \frac{\langle A\rangle_\phi}{\|\phi\|^2} = \frac{\int_{\mathbb{R}^3} \overline{\phi} A\phi}{\int_{\mathbb{R}^3} |\phi|^2}\,.$$

In particular, if $\phi \in \mathcal{H}$ is a normalized state (of unit norm), then

$$\overline{a}_\phi = \langle A\rangle_\phi\,.$$

The third postulate has the following statistical interpretation: we know from Postulate I that the measure of A in the state ϕ must result in one of the eigenvalues of A. If we repeat the measurement in many states, all of them identical to ϕ, two cases are possible: if ϕ is one of the eigenfunctions of A, corresponding to the eigenvalue $a \in \mathbb{R}$, the result will be a in all the measurements (by Postulate II); but if ϕ is not an eigenfunction, we will obtain a sample of eigenvalues, and Postulate III establishes what the mean of that sample must be.

Along with these postulates of mathematical character, there are others of a more philosophical kind, that serve the purpose of making contact with physical computations. The first one of these is the *Complementarity Principle* due to Bohr, stating the impossibility of avoiding the wave-particle duality introduced by De Broglie.

Postulate A The complete knowledge of the characteristics of a quantum system require a description of its particle-like and wave-like properties. On the other hand, it is not possible to perform a *single* experiment where both kind of properties are determined *simultaneously*.

Postulate B Quantum Mechanics, when applied to macroscopic systems in which the perturbations accompanying the measurements are negligible, must reproduce the results of Classical Mechanics.

Postulate C Given a wave function $\phi \in \mathcal{H}$, representing the state of a physical system, the function

$$\rho(t, \mathbf{x}) = |\phi|^2(t, \mathbf{x}) = \overline{\phi}(t, \mathbf{x})\phi(t, \mathbf{x})$$

represents the *probability density of localizing the system* in space. Thus, if $D \subset \mathbb{R}^3$, the probability that the system occupies some position in D at instant $t \in \mathbb{R}$ is given by

$$P_D(t) = \int_D |\phi|^2(t, \mathbf{x})\, d\mathbf{x}\,.$$

Moreover, if the system is in state $\phi \in \mathcal{H}$, the probability amplitude of finding it in the state $\varphi \in \mathcal{H}$ is $\langle \phi, \varphi \rangle$; the probability itself is $|\langle \phi, \varphi \rangle|^2$, where it is supposed that the states are normalized.

Postulate B is the *Principle of Correspondence*, and Postulate C is known as *Born's interpretation* of the (squared modulus of the) wave function.

2.3 Wave equations: Schrodinger and Klein-Gordon

We have seen that Von Neumann formalism assign self-adjoint operators on a Hilbert space to observable properties of a physical system, but still we do not have a single example of this correspondence. We will employ the postulates of the previous section to construct the operators corresponding to position and momentum and, in doing so, we will also derive the basic evolution equations of the theory.

Remark 9. We will use the following notation:

$$\begin{aligned} \hbar &= \frac{h}{2\pi} \text{ (Planck's constant)} \\ \omega &= 2\pi\nu \text{ (angular frequency)} \\ k &= \frac{2\pi}{\lambda} \text{ (wave number)}\,. \end{aligned}$$

Also, for simplicity, we will work in one spatial dimension.

An elementary wave is described in terms of these quantities by

$$\phi(t, x) = e^{i(kx - \omega t)}\,.$$

The angular frequency and the wave number are clearly functionally dependent (there must be a relation between the number of cycles in a second and the number of complete waves in a length of 2π) and the function

$$\omega = \omega(k)$$

is called the *dispersion relation*, a characteristic of each wave. In the case in which $\phi(t, x)$ represents a free non-relativistic quantum particle, we can find its dispersion relation from the De Broglie formulas: starting from the expression of the energy

$$E = \frac{p^2}{2m} \tag{14}$$

and using $E = h\nu$, $p = h/\lambda$ in the forms $E = \hbar\omega$, $p = \hbar k$, respectively, we find $\hbar\omega = \hbar^2 k^2/2m$, giving the dispersion relation

$$\omega(k) = \frac{\hbar}{2m} k^2 \,. \tag{15}$$

The description of a quantum free particle by a single wave, however, is not a good one. The reason is that for a single wave we have the *phase velocity*, given by the quotient $v_f = \omega/k$. In our case,

$$v_f = \frac{1}{k}\frac{\hbar}{2m}k^2 = \frac{\hbar k}{2m} = \frac{p}{2m}\,,$$

and this does not coincide with the classical velocity of the particle, p/m. Thus, we are not conforming to the Correspondence Principle. The solution is to consider *wave packets*: superpositions of single waves with different wave numbers, such as

$$\psi(t, x) = \int a(k) e^{i(kx-\omega t)}\, dk \,.$$

For such a packet, each wave has a distinctive dispersion relation $\omega = \omega(k)$, and thus each wave in the packet evolves at a different speed. In this case, the velocity to be attached to the wave as a whole is the *group velocity*

$$v_g = \frac{d\omega}{dk} \,.$$

Now we find

$$v_g = \frac{d\omega}{dk} = \frac{d}{dk}\left(\frac{\hbar k^2}{2m}\right) = \frac{\hbar k}{m} = \frac{p}{m} = v\,,$$

which is the particle velocity.

Now that we know the wave packets are the correct way of describing particles in Quantum Mechanics, we can attack the problem of finding the evolution equation for them. The most obvious thing is to try a wave equation type; thus, we compute:

$$\frac{\partial \psi}{\partial t} = -\int i\omega a(k) e^{i(kx-\omega t)}\, dk$$

and

$$\frac{\partial^2 \psi]}{\partial x^2} = -\int k^2 a(k) e^{i(kx-\omega t)}\, \mathring{a}k \,.$$

To relate these partial derivatives, we use the dispersion relation (15), which tells us

$$\int \left(\hbar\omega - \frac{\hbar^2 k^2}{2m} \right) a(k) e^{kx-\omega t} \, dk .$$

This expression gives us the clue to add the needed factors, and we arrive at the famous *Schrodinger equation*:

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} . \tag{16}$$

Schrodinger's equation joint with the Correspondence Principle will be enough to determine the operators of position and momentum. As a consequence of Born's postulate, the probability density for the results of measuring a position is given by $|\psi|^2$, so (in the statistical sense) the expected value for the position is that imposed by Postulate III (assuming ψ normalized):

$$\overline{x} = \int x|\psi(x)|^2 \, dx = \int \overline{\psi}(x) x \psi(x) \, dx = \langle X \rangle_\psi ,$$

a relation that leads us to take the operator

$$X\psi(x) = x\psi(x) \tag{17}$$

as the observable representing position in Quantum Mechanics.

To find the operator associated to momentum, we will use the Principle of Correspondence: the derivative with respect to time of $\langle X \rangle_\psi$ must be the average momentum divided by mass (that is, the velocity), $\langle P \rangle_\psi / m = \langle P/m \rangle_\psi$. But we can compute

$$\frac{d}{dt} \langle X \rangle_\psi = \frac{d}{dt} \int \overline{\psi}(x) x \psi(x) \, dx = \frac{d}{dt} \int \left(\frac{\partial \overline{\psi}}{\partial t}(x) x \psi(x) + \overline{\psi}(x) x \frac{\partial \psi}{\partial t}(x) \right) dx ,$$

and using (16), integration by parts and the fact that each square integrable function has a representative of compact support,

$$\begin{aligned} \frac{d}{dt} \langle X \rangle_\psi &= \frac{1}{i\hbar} \int \left(\frac{\hbar^2}{2m} \frac{\partial^2 \overline{\psi}}{\partial x^2} x\psi + \overline{\psi} x \left(-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} \right) \right) dx \\ &= \int \overline{\psi} \left(-\frac{i\hbar}{m} \frac{\partial \psi}{\partial x} \right) dx \\ &= \left\langle -\frac{i\hbar}{m} \frac{\partial}{\partial x} \right\rangle_\psi . \end{aligned}$$

Therefore, the momentum operator must be taken as

$$P = -i\hbar \frac{\partial}{\partial x} . \tag{18}$$

Remark 10. The asignments (17) and (18) are called the *Dirac (canonical) quantization rules*. Notice that these operators constitute a Fourier conjugate pair. Also, we have the following *commutation relations* between position and momentum

$$[P, X] = -i\hbar \mathrm{Id} , \tag{19}$$

a result in accordance with the considerations we made in Remark 8.

Remark 11. A reasoning similar to the one just presented leads to the *energy operator*,

$$E = i\hbar \frac{\partial}{\partial t}\,. \tag{20}$$

In arriving at the Schrodinger equation (16) que departed from the classical expression of the energy (14). Now that we know the basic observables of the theory, we can turn the argument inside out, and construct a *relativistic* wave equation using the Dirac quantization but applied to the relativistic energy (9):

$$E^2 - \|\mathbf{p}\|^2 = m^2\,.$$

Doing so, we get the equation

$$-\hbar^2 \frac{\partial \psi}{\partial t} + \hbar^2 \frac{\partial^2 \psi}{\partial x^2} = m^2 \psi$$

that can be expressed as

$$(\Box + m^2)\psi = 0\,, \tag{21}$$

where $\Box = \partial_\mu \eta_{\mu\nu} \partial_\nu$ is the Dalembertian operator in 4−dimensions, and $\partial_\mu \equiv (\partial_0, \partial_i)$.

Equation (21) is the *Klein-Gordon equation*, a relativisitic quantum wave equation that describes particles without spin.

2.4 The harmonic oscillator

In Classical Mechanics, the harmonic oscillator is the system defined by the force $F(x) = -mw^2 x$. That force can be derived from the potential

$$V(x) = -\int^x F(s)\,ds = \frac{mw^2}{2}\,x^2\,.$$

Thus, the classical Hamiltonian for this system (its total energy) is

$$H = \frac{p^2}{2m} + \frac{mw^2}{2} x^2\,.$$

The corresponding quantum Hamiltonian (which will be denoted with the same letter) is found with the aid of the Correspondence Principle:

$$H = \frac{1}{2m} P^2 + \frac{1}{2} m w^2 X^2\,.$$

Writing for the wave equation $\psi(t,x) = T(t)\phi(x)$, it is readily found (by separation of variables) that

$$T(t) = e^{-i\frac{E}{\hbar}t}\,,$$

and we are left with the *time-independent* Schrodinger equation for $\phi(x)$:

$$-\frac{\hbar^2}{2m}\frac{d^2\phi}{dx^2} + \left(\frac{1}{2} m w^2 x^2 - E\right)\phi(x) = 0\,.$$

This equation can be solved with the series techniques available for second-order differential equations with analytic coefficients (such as the Frobenius

method), but here we will follow an idea due to Dirac, related to the theory of representations of Lie algebras. First of all, let us make a rescaling

$$p = \frac{1}{\sqrt{mw\hbar}} P \text{ and } q = \sqrt{\frac{mw}{\hbar}} X,$$

so the Hamiltonian reads now

$$H = \frac{\hbar w}{2}(p^2 + q^2). \tag{22}$$

Let us note that the commutation relation (19) now is given by

$$[p, q] = -i\mathrm{Id}.$$

Following Dirac, the next step consists in introducing the operators[†]

$$\begin{aligned} a &\doteq \frac{1}{\sqrt{2}}(q + ip) \\ a^\dagger &\doteq \frac{1}{\sqrt{2}}(q - ip). \end{aligned} \tag{23}$$

Substituting in (22) we get

$$H = \hbar w \frac{1}{2}(a^\dagger a + a a^\dagger),$$

and the new operators satisfy the commutation relations

$$\begin{aligned} [a, a] = 0 = [a^\dagger, a^\dagger] \\ [a, a^\dagger] = \mathrm{Id}. \end{aligned} \tag{24}$$

For reasons that will become clear later, it is also convenient to introduce the so-called *number operator*

$$N = a^\dagger a, \tag{25}$$

so we finally arrive at the expression for the Hamiltonian

$$H = \hbar w \left(N + \frac{1}{2} \right).$$

To study the spectrum of the harmonic oscillator (or its Hamiltonian H), basically tantamount to study the spectrum of N, and this we do now. We will make use only of the algebraic properties (24) (this was the genius of Dirac).

Proposition 2. *Let ν be an eigenvalue of N with associated eigenvector $|\nu\rangle$. Then, ν is a positive real number, $\nu \in \mathbb{R}^+$.*

Proof. From the definition (25) is clear that $N = N^\dagger$, so its eigenvalues are real. If $|\nu\rangle$ is an eigenvector corresponding to ν, let us call $|\mu\rangle = a|\nu\rangle$. Then:

$$0 \leqslant \langle \mu|\mu\rangle = \langle \nu|a^\dagger a|\nu\rangle = \langle \nu|N|\nu\rangle = \nu \langle \nu|\nu\rangle .$$

As $\langle \nu|\nu\rangle \, \|\nu\|^2 \geqslant 0$, we get $\nu \geqslant 0$. □

† The notation is due to the fact that a and $a^\dagger$ are formally adjoint one to each other, on the common domain on which p and q are self-adjoint operators.

In exercise 7 we prove that 0 is an eigenvalue of N and, moreover, it is non-degenerate. Then, the following result makes sense.

Proposition 3. *If* $|0\rangle$ *denotes the state described by the eigenvector corresponding to* $\nu = 0$*, then* $a|0\rangle = 0$*.*

Proof. Again making $|\mu\rangle = a|0\rangle$, a quick computation shows that

$$\langle\mu|\mu\rangle = \langle 0a|a^\dagger a|0\rangle = \langle 0|N|0\rangle = 0\,\langle 0|0\rangle = 0\,,$$

and the non-degeneracy of the scalar product implies $|\mu\rangle = 0$. □

Proposition 4. *If* $|\nu\rangle$ *is an eigenvector of* N *corresponding to the eigenvalue* ν*, then,* $a^\dagger|\nu\rangle$ *is an eigenvector of* N *corresponding to the eigenvalue* $\nu + 1$*.*

Proof. We will use the commutation relation $[N, a^\dagger] = a^\dagger$ in the form

$$Na^\dagger = a^\dagger(N+1)\,.$$

Again putting $|\mu\rangle = a^\dagger|\nu\rangle$ and using the hypothesis, we find

$$N|\mu\rangle = Na^\dagger|\nu\rangle = a^\dagger(N+1)|\nu\rangle = (\nu+1)a^\dagger|\nu\rangle = (\nu+1)|\mu\rangle\ ,$$

which is the statement. □

Proposition 5. *If* $|\nu\rangle \neq 0$ *is an eigenvector of* N *corresponding to the eigenvalue* ν*, then,* $a|\nu\rangle$ *is an eigenvector of* N *corresponding to the eigenvalue* $\nu - 1$*.*

Proof. The proof is the same as that of the preceding Proposition, only that this time we will use the commutation relation $[N, a] = -a$ in the form $Na = a(N-1)$. Putting $|\mu\rangle = a|\nu\rangle$ and computing, we get

$$N|\mu\rangle = Na|\nu\rangle = a(N-1)|\nu\rangle = (\nu-1)a|\nu\rangle = (\nu-1)|\mu\rangle\ ,$$

which is what we wanted to prove. □

The next result explains the name 'number operator' for N.

Theorem 2. *The spectrum of* N *is* $\sigma(N) = \{0\} \cup \mathbb{N}$*.*

Proof. Let us suppose first that ν is *not* an integer. By the preceding results, the vectors $a|\nu\rangle, a^2|\nu\rangle, a^3|\nu\rangle, \ldots$, are all eigenvectors of N with respective eigenvalues $\nu-1, \nu-1, \nu-1, \ldots$ Let us put $|\mu\rangle = a|\nu - s\rangle$ for a fixed $s \in \mathbb{N}$, and compute

$$\langle\mu|\mu\rangle = (\nu - s)\,\langle\nu - s|\nu - s\rangle\ .$$

Because $|\mu\rangle$ is a vector in a Hilbert space, its squared norm satisfies $\langle\mu|\mu\rangle \geqslant 0$. As $|\nu - s\rangle$ is a vector too, $\langle\nu - s|\nu - s\rangle \geqslant 0$, and these conditions force $\nu - s \geqslant 0$, with $s \in \mathbb{N}$ fixed but arbitrary. If ν is not an integer, there is no way in which this condition can be satisfied for s big enough. Thus, in case ν is not an integer we will be contradicting Proposition 2 above.

On the other hand, if $\nu \in \mathbb{N} \cup \{0\}$ the sequence of eigenvalues, according to the preceding Proposition, ends in 0 before negative eigenvalues appear, due to $a|0\rangle = 0$. In this case, there is no problem, and ν is an eigenvalue (we will construct later on its associated eigenvector explicitly, in Exercise 8). □

The eigenvector $|0\rangle$, corresponding to the eigenvalue 0, is called the *base state* or *vacuum state* of the harmonic oscillator. If we impose the normalization

$$\langle n|n'\rangle = \delta_{nn'}$$

(where we use $|n\rangle$ instead of $|\nu\rangle$ to emphasize that the eigenvalues of N are in $\{0\} \cup \mathbb{N}$), the results we already have allow us to conclude

$$a^\dagger |n\rangle = c_n |n+1\rangle \ ,$$

and we can compute the normalization constant as follows:

$$\begin{aligned} 1 = \langle n+1|n+1\rangle &= \frac{1}{|c_n|^2} \langle n|aa^\dagger|n\rangle \\ &= \frac{1}{|c_n|^2} \langle n|N-1|n\rangle \\ &= \frac{n-1}{|c_n|^2} \langle n|n\rangle \\ &= \frac{n-1}{|c_n|^2} \end{aligned}$$

so $c_n = \sqrt{n-1}$, and we can put

$$a^\dagger |n\rangle = \sqrt{n+1}\,|n+1\rangle \ . \tag{26}$$

Analogously, we can write $a|n\rangle = k_n |n-1\rangle$, and a similar computation (left to the reader) will determine the constant k_n to give

$$a|n\rangle = \sqrt{n}\,|n-1\rangle \ . \tag{27}$$

With the aid of (26) and (27), we can find the following sequence of eigenvectors of N:

$$\begin{aligned} &|0\rangle \\ &|1\rangle = a^\dagger |0\rangle \\ &|2\rangle = \frac{1}{\sqrt{2!}} \left(a^\dagger\right)^2 |0\rangle \\ &\vdots \\ &|k\rangle = \frac{1}{\sqrt{k!}} \left(a^\dagger\right)^k |0\rangle \ . \end{aligned} \tag{28}$$

Due to this property, $a^\dagger$ is called the *creation operator* (as it generates the $|k\rangle$ state from the vacuum by a reiterated action). Analogously, we can form the sequence

$$\begin{aligned} |0\rangle &= a|1\rangle \\ |1\rangle &= \frac{1}{\sqrt{2}} a|2\rangle \\ |2\rangle &= \frac{1}{\sqrt{3}} a|3\rangle \\ &\vdots \\ |k\rangle &= \frac{1}{\sqrt{k}} a|k+1\rangle \ . \end{aligned} \tag{29}$$

which justifies the name *annihilation operator* for a.

2.5 Exercises

1. Prove that the operator corresponding to the observable energy is

$$E = i\hbar \frac{\partial}{\partial t}.$$

2. Starting from the classical expression for the energy in the presence of a potential $V(x)$,

$$E = \frac{p^2}{2m} + V(x),$$

prove that the Schrodinger equation is

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V\psi.$$

In Physics, the classical function that gives the total energy of a system is also called the *Hamiltonian*. Thus, the quantum mechanical Hamiltonian is the operator

$$H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V,$$

and the Schrodinger equation can be written as

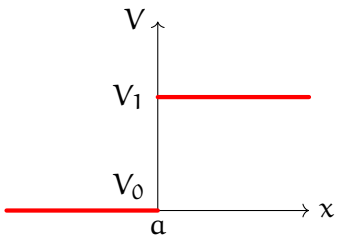
$$H\psi = E\psi,$$

which can be interpreted as the problem of finding the eigenvalues of H (physically they represent the allowed energies for the system).

3. By applying both sides to a wave function, prove the commutation relation (19).

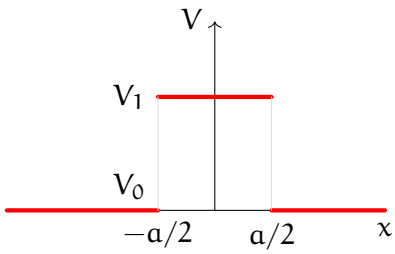
4. Solve the Schrodinger equation for the potential step:

$$V(x) \doteq \begin{cases} V_0 & \text{if } x < a \\ V_1 & \text{if } x > a. \end{cases}$$



5. Solve the Schrodinger equation for the potential square barrier:

$$V(x) \doteq \begin{cases} V_0 & \text{si } x < -a/2 \\ V_1 & \text{si } -a/2 < x < a/2 \\ V_0 & \text{si } x > a/2. \end{cases}$$



In particular, determine the probability of crossing the barrier that a particle with energy $E < V_1$ has. This phenomenon is the celebrated *tunnel effect*.

6. Prove that the eigenvalues of the number operator N have the following property: if the eigenvalue n is non-degenerate, then $n+1$ it is also non-degenerate.

7. Exercise 6 can be applied to conclude that each eigenvalue of N is non-degenerate, it only remains to prove that the eigenspace corresponding to $n = 0$ is one-dimensional. Do this by explicitly solving the equation $a|0\rangle = 0$, that is, find a wave function $\phi_0(x)$ such that $a\phi_0(x) = 0$ (so $N\phi_0 = 0$).

8. Starting from the eigenvector ϕ_0 constructed in the preceding exercise, give an explicit formula to compute the eigenvector ϕ_n of the harmonic oscillator corresponding to n.

REFERENCES

Three modern references are:

[1] J. Dimock: *Quantum Mechanics and Quantum Field Theory: A Mathematical Primer*. Cambridge UP, 2011.

[2] L. D. Faddeev and O. A.Yakubovskii: *Lectures on Quantum Mechanics for Mathematics Students*. AMS Publishing, 2009.

[3] J. Franklin: *Classical Field Theory*. Cambridge University Press, 2017.

However, the old texts are really worth reading for a newcomer. Landau and Lifshitz wrote a superb exposition of the classical theory of electromagnetic and gravitational fields, with lots of examples and detailed calculations. The book by Itzykson and Zuber is the best ever written on QFT, while that of Mandl is brief and goes directly to the point, at a more elementary level. Our presentation of Quantum Mechanics follows that of Matthews.

[4] C. Itzykson and J. B. Zuber: *Quantum Field Theory*. Dover, 2005.

[5] L. D. Landau and E. M. Lifshitz: *The Classical Theory of Fields*. Pergamon Press, 1971. Available on line at the URL `https://archive.org/details/TheClassicalTheoryOfFields`.

[6] F. Mandl: *Quantum Field Theory*. Interscience Publishers, 1959. Available on line at the URL `archive.org/details/IntroductionToQuantumFieldTheory`.

[7] P. T. Matthews: *Introduction to Quantum Mechanics*. 2nd Ed. McGraw-Hill, 1968.

Also, there are a lot of resources that can be found in the Internet. Readers of the present notes should be able to follow a full course such as the excellent one by D. Tong referenced below.

[8] D. Tong: *Quantum Field Theory*. University of Cambridge Part III Mathematical Tripos, 2006–07. Available on line at the URL `http://www.damtp.cam.ac.uk/user/tong/qft.html`.