

Chapter 1

Introduction



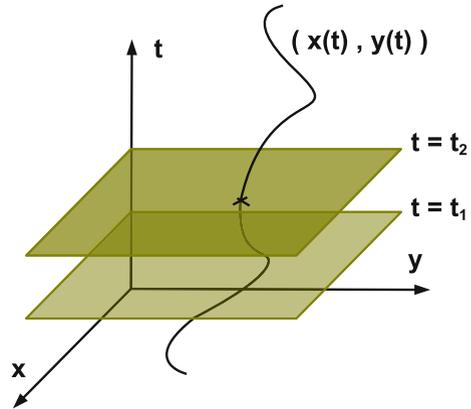
This chapter briefly reviews the Euclidean geometry, Galilean transformations, the Lagrangian formalism, and Newton's gravity. While the reader is supposed to be already familiar with all these concepts, it is convenient to summarize them here because they will be used or generalized in the next chapters for the theories of special and general relativity. We end the chapter pointing out the inconsistency between Galilean transformations and Maxwell's equations and how this issue led to the theory of special relativity between the end of the 19th century and the beginning of the 20th century.

1.1 Special Principle of Relativity

Let us consider the motion of a point-like particle in an n -dimensional space. If we want to describe such a physical system, we intuitively need $n + 1$ variables, namely n space coordinates to describe the position of the particle in space at a certain value of the temporal coordinate (see Fig. 1.1). In order to assign the n space coordinates to the particle, we need to measure the distance and the direction of the particle from a certain reference point with a standard rod. The temporal coordinate is determined by measuring the time interval with respect to a certain reference time with a standard clock. The choices of the reference point and of the reference time, as well as those of the standard rod and of the standard clock, correspond to the choice of a certain observer, namely of a particular *reference frame*. Two natural questions are:

1. Should we choose some particular reference frame to describe the motion of the point-like particle? In other words, is there any preferred observer/reference frame, or any class of preferred observers/reference frames, or are the laws of physics independent of such a choice?

Fig. 1.1 Motion of a point-like particle in a 2-dimensional space. x and y are the space coordinates and t is the time. The trajectory of the particle is described by the curve $(x(t), y(t))$



2. How are the physical quantities measured in a certain reference frame related to the same physical quantities measured in another reference frame?

Galileo Galilei was the first, in the 17th century, to discuss the issue of the choice of reference frame to describe physical phenomena. From simple observations, we can realize that there is a certain class of observers that turns out to be particularly suitable to describe physical phenomena. This is the class of inertial observers (or inertial reference frames).

Inertial Reference Frame. An *inertial reference frame* is a reference frame in which the motion of a body not subject to forces either remains at rest or continues to move at a constant speed in a straight line.

While it is possible to describe physical phenomena even in non-inertial reference frames, namely in the reference frames not belonging to the class of inertial reference frames, the description is more complicated. In particular, it is usually necessary to introduce some (reference frame-dependent) corrections to the laws of physics.

Note that, strictly speaking, inertial reference frames do not exist in Nature, as in the Universe there are long-range forces that cannot be screened. Nevertheless, we can usually find reference frames that well approximate inertial ones.

With the concept of inertial reference frame we can introduce the Special Principle of Relativity.

Special Principle of Relativity. The laws of physics are the same in all inertial reference frames.

As a principle, the Special Principle of Relativity cannot be proved by theoretical arguments, but only confirmed (or disproved) by experiments. Current experiments

and observational data support this principle. However, there are still attempts today to test the Special Principle of Relativity with higher and higher precision or in different environments, as well as theoretical models in which this principle can be violated at some level.

1.2 Euclidean Space

Let us consider a 3-dimensional space with the system of Cartesian coordinates (x, y, z) . Such a space can be “identified” with \mathbb{R}^3 , because every point of the space can be characterized by three real numbers, which are the values of the coordinates (x, y, z) .

The infinitesimal distance between the point $\mathbf{x}_A = (x_A, y_A, z_A)$ and the point $\mathbf{x}_B = (x_A + dx, y_A + dy, z_A + dz)$ is the square root of

$$dl^2 = dx^2 + dy^2 + dz^2. \quad (1.1)$$

dl is called the *line element* and Eq. (1.1) simply follows from Pythagoras’s theorem.

It is convenient to introduce the notation (x^1, x^2, x^3) to denote the coordinates of the space. In the case of Cartesian coordinates, we have $x^1 = x$, $x^2 = y$, and $x^3 = z$. Now Eq. (1.1) can be written in a more compact form as

$$dl^2 = \delta_{ij} dx^i dx^j, \quad (1.2)$$

where δ_{ij} is the Kronecker delta and we have adopted the Einstein convention of summation over repeated indices; that is,

$$\delta_{ij} dx^i dx^j \equiv \sum_{i,j=1}^3 \delta_{ij} dx^i dx^j. \quad (1.3)$$

δ_{ij} is the *Euclidean metric* and can be written as the matrix

$$||\delta_{ij}|| = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (1.4)$$

Our discussion can be easily extended to any n -dimensional space. The n -dimensional *Euclidean space* is \mathbb{R}^n in which the square of the line element is¹

¹From the mathematical point of view, the n -dimensional Euclidean space is the differentiable manifold \mathbb{R}^n equipped with the Euclidean metric δ_{ij} (See Appendix C).

$$dl^2 = \delta_{ij} dx^i dx^j, \quad (1.5)$$

where now i and j run from 1 to n .

Note that the infinitesimal distance between two points is independent of the coordinate system. The line element is thus an *invariant*; that is, it does not change with a change of coordinates. For an arbitrary coordinate system, we write the square of the line element as

$$dl^2 = g_{ij} dx^i dx^j, \quad (1.6)$$

where g_{ij} is called the *metric tensor* and, in general, is not δ_{ij} . Generally speaking, if we move from the coordinate system (x^1, x^2, x^3) to the coordinate system (x'^1, x'^2, x'^3) we have

$$dx^i \rightarrow dx'^i = \frac{\partial x'^i}{\partial x^j} dx^j, \quad (1.7)$$

and therefore

$$g_{ij} dx^i dx^j = g'_{ij} dx'^i dx'^j = g'_{ij} \frac{\partial x'^i}{\partial x^m} dx^m \frac{\partial x'^j}{\partial x^n} dx^n. \quad (1.8)$$

We see that

$$g_{mn} = \frac{\partial x'^i}{\partial x^m} \frac{\partial x'^j}{\partial x^n} g'_{ij}. \quad (1.9)$$

We multiply both sides of this expression by $\partial x^m / \partial x'^p$ and $\partial x^n / \partial x'^q$, and we sum over repeated indices

$$\frac{\partial x^m}{\partial x'^p} \frac{\partial x^n}{\partial x'^q} g_{mn} = \frac{\partial x^m}{\partial x'^p} \frac{\partial x^n}{\partial x'^q} \frac{\partial x'^i}{\partial x^m} \frac{\partial x'^j}{\partial x^n} g'_{ij} = \delta_p^i \delta_q^j g'_{ij} = g'_{pq}. \quad (1.10)$$

The metric tensor thus transforms as

$$g_{ij} \rightarrow g'_{ij} = \frac{\partial x^m}{\partial x'^j} \frac{\partial x^n}{\partial x'^i} g_{mn}. \quad (1.11)$$

As an example, we can consider the spherical coordinates (r, θ, ϕ) . The relation between Cartesian and spherical coordinates is

$$\begin{aligned} x &= r \sin \theta \cos \phi, \\ y &= r \sin \theta \sin \phi, \\ z &= r \cos \theta, \end{aligned} \quad (1.12)$$

with inverse

$$\begin{aligned}
 r &= \sqrt{x^2 + y^2 + z^2}, \\
 \theta &= \arccos\left(\frac{z}{\sqrt{x^2 + y^2 + z^2}}\right), \\
 \phi &= \arctan\left(\frac{y}{x}\right).
 \end{aligned} \tag{1.13}$$

It is straightforward to apply Eq. (1.11) and see that in spherical coordinates the square of the line element is

$$dl^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2, \tag{1.14}$$

and therefore the corresponding metric tensor reads

$$\|g_{ij}\| = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}. \tag{1.15}$$

With the concept of a line element, we can measure the length of a curve. In the 3-dimensional space with Cartesian coordinates, a curve is a continuous function $\Gamma: t \in [t_1, t_2] \subset \mathbb{R} \rightarrow \mathbb{R}^3$. The points of the curve have the coordinates

$$\mathbf{x}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}. \tag{1.16}$$

The length of the curve is

$$\ell = \int_{\Gamma} dl = \int_{t_1}^{t_2} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} dt, \tag{1.17}$$

where here the dot indicates the derivative with respect to the parameter t . The length of a curve between two points of the space is an invariant as well.

As an example of the length of a curve, let us consider a circle in \mathbb{R}^2 . The points of the circle have Cartesian coordinates

$$\mathbf{x}(t) = \begin{pmatrix} R \cos t \\ R \sin t \end{pmatrix}, \tag{1.18}$$

where R is the radius of the circle and $t \in [0, 2\pi)$. The length of the curve is

$$\ell = \int_0^{2\pi} \sqrt{R^2 \sin^2 t + R^2 \cos^2 t} dt = \int_0^{2\pi} R dt = 2\pi R. \tag{1.19}$$

1.3 Scalars, Vectors, and Tensors

A *scalar* ϕ is a quantity that does not change if we change coordinates: under the coordinate transformation $x^i \rightarrow x'^i$, we have

$$\phi \rightarrow \phi' = \phi. \quad (1.20)$$

For example, the line element dl is a scalar.

A *vector* is, strictly speaking, an element of a vector space, which is a set of objects in which we can define two operations (addition and multiplication) satisfying certain axioms. The reader is presumably already familiar with the concept of vectors, but more details can be found in Appendix A.2. For instance, the infinitesimal displacement between two nearby points of the space,

$$d\mathbf{x} = (dx^1, dx^2, \dots, dx^n), \quad (1.21)$$

is a vector. Note that the use of upper and lower indices in the previous section was not accidental. Upper indices are employed for the components of vectors, which transform as

$$V^i \rightarrow V'^i = \frac{\partial x'^i}{\partial x^j} V^j, \quad (1.22)$$

for the change of coordinates² $x^i \rightarrow x'^i$.

Lower indices are used for the components of a *dual vector* (also called *cotangent vector* or *co-vector*), which transform as

$$V_i \rightarrow V'_i = \frac{\partial x^j}{\partial x'^i} V_j. \quad (1.23)$$

In this book, the dual vector of the vector $\mathbf{V} = (V^1, V^2, \dots, V^n)$ is indicated as \mathbf{V}^* and defined as the object with components

$$V_i \equiv g_{ij} V^j. \quad (1.24)$$

The dual vector can thus be seen as a function that requires as input a vector (let us write its components as W^i) and provides as output a real number

$$V_i(W^i) = g_{ij} W^i V^j. \quad (1.25)$$

²The coordinates of a space, $\{x^i\}$, are not the components of a vector even if they have upper indices. Indeed they do not transform with the rule (1.22) in general. For example, this is easy to check with the transformations between Cartesian and spherical coordinates in Eqs. (1.12) and (1.13). We write the space coordinates with upper indices because it is common to do so and we have to write the indices somewhere.

Note that a quantity like $V_i W^i$ is a scalar, namely it is invariant under a coordinate transformation

$$V_i W^i \rightarrow V'_i W'^i = \frac{\partial x^j}{\partial x'^i} V_j \frac{\partial x'^i}{\partial x^k} W^k = \delta_k^j V_j W^k = V_j W^j. \quad (1.26)$$

We say that upper indices are “lowered” by the metric tensor g_{ij} , as shown in Eq. (1.24), and lower indices are “raised” by the inverse of the metric tensor g^{ij}

$$V^i = g^{ij} V_j = g^{ij} g_{jm} V^m = \delta_m^i V^m = V^i, \quad (1.27)$$

where $g^{ij} g_{jm} = \delta_m^i$ by definition. When we use Cartesian coordinates, we have the Euclidean metric δ_{ij} and its action on a vector is trivial: if we have the vector $\mathbf{V} = (V^x, V^y, V^z)$, the dual vector is

$$\mathbf{V}^* = \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix} = \begin{pmatrix} V^x \\ V^y \\ V^z \end{pmatrix}. \quad (1.28)$$

However, this is not the general case. If we consider spherical coordinates, the dual vector of the vector $\mathbf{V} = (V^r, V^\theta, V^\phi)$ is

$$\mathbf{V}^* = \begin{pmatrix} V_r \\ V_\theta \\ V_\phi \end{pmatrix} = \begin{pmatrix} V^r \\ r^2 V^\theta \\ r^2 \sin^2 \theta V^\phi \end{pmatrix}. \quad (1.29)$$

Tensors are the generalization of vectors and dual vectors. They are multi-index objects. An example is the metric tensor g_{ij} . A tensor of type (r, s) and of order $r + s$ has r upper indices and s lower indices. The transformation rule for the components of a tensor is

$$T_{j_1 j_2 \dots j_s}^{i_1 i_2 \dots i_r} \rightarrow T'_{j'_1 j'_2 \dots j'_s}^{i'_1 i'_2 \dots i'_r} = \frac{\partial x'^{i'_1}}{\partial x^{p_1}} \frac{\partial x'^{i'_2}}{\partial x^{p_2}} \dots \frac{\partial x'^{i'_r}}{\partial x^{p_r}} \frac{\partial x^{q_1}}{\partial x'^{j'_1}} \frac{\partial x^{q_2}}{\partial x'^{j'_2}} \dots \frac{\partial x^{q_s}}{\partial x'^{j'_s}} T_{q_1 q_2 \dots q_s}^{p_1 p_2 \dots p_r}. \quad (1.30)$$

A scalar is a tensor of type $(0,0)$, a vector is a tensor of type $(1,0)$, and a dual vector is a tensor of type $(0,1)$. Upper indices can be lowered with g_{ij} , lower indices can be raised with g^{ij} . Some examples are³

$$T^{ijk} = g^{il} g^{jm} g^{kn} T_{lmn}, \quad T^i_{jk} = g^{il} g_{jm} T_l^m{}_k, \quad \dots \quad (1.31)$$

³Note that, in general, the order of the indices is important. For instance, if we have the tensor A^{ab} and we lower the index a , we should write $A_a{}^b$. If we lower the index b , we should write $A^a{}_b$. The index a must remain the first from the left. If A^{ab} is a symmetric tensor, i.e. $A^{ab} = A^{ba}$, we have also $A_a{}^b = A^b{}_a$, the order does not matter, and we can simplify the notation writing A^a_b .

If we raise an index of the metric tensor, we get the Kronecker delta, $g^{ij}g_{jm} = g^i_m = \delta^i_m$.

When we sum over repeated upper and lower indices, we reduce the order of the tensor. For instance, if we have the tensor of type (2,1) with components T^{ij}_k and we “contract” the indices i and k , we get the vector

$$V^j = T^{ij}_i. \quad (1.32)$$

Indeed, if we consider the transformation of coordinates $x^i \rightarrow x'^i$, we have

$$\begin{aligned} V^j \rightarrow V'^j &= T'^{ij}_i = \frac{\partial x'^i}{\partial x^l} \frac{\partial x'^j}{\partial x^m} \frac{\partial x^n}{\partial x'^i} T^{lm}_n = \delta^n_l \frac{\partial x'^j}{\partial x^m} T^{lm}_n = \frac{\partial x'^j}{\partial x^m} T^{lm}_l \\ &= \frac{\partial x'^j}{\partial x^m} V^m, \end{aligned} \quad (1.33)$$

and V^j transforms as a vector. If we have a tensor of type (r, s) and of order $r + s$ and we contract $2t$ indices, the new tensor is of type $(r - t, s - t)$ and of order $r + s - 2t$. We can also contract indices between two different tensors. For instance, if we have the vector with components V^i and the dual vector W_j and we contract over their indices, we get the scalar $V^i W_i$, which is a number that does not change under a change of coordinates.

If some physical quantity is described by a tensor of type (r, s) at every point of the space, we have a *tensor field* of type (r, s) . A scalar field is thus a function of the form $\phi = \phi(x^1, x^2, \dots, x^n)$. A vector field has the form

$$\mathbf{V} = \mathbf{V}(x^1, x^2, \dots, x^n) = \begin{pmatrix} V^1(x^1, x^2, \dots, x^n) \\ V^2(x^1, x^2, \dots, x^n) \\ \vdots \\ V^n(x^1, x^2, \dots, x^n) \end{pmatrix}, \quad (1.34)$$

that is, every component is a function of the space coordinates.

A more rigorous definition of vectors, dual vectors, and tensors can be found in Appendices A and C.

1.4 Galilean Transformations

Let us consider an inertial reference frame with the Cartesian coordinates $\mathbf{x} = (x, y, z)$ and the time t and another inertial reference frame with the Cartesian coordinates $\mathbf{x}' = (x', y', z')$ and the time t' . Let us also assume that the second reference frame is moving with constant velocity \mathbf{v} with respect to the former. If the two Cartesian coordinates coincide at the time $t = t' = 0$, the transformation connecting the two reference frames is

$$\begin{aligned}\mathbf{x} &\rightarrow \mathbf{x}' = \mathbf{x} - \mathbf{v}t, \\ t &\rightarrow t' = t.\end{aligned}\tag{1.35}$$

If, for instance, $\mathbf{v} = (v, 0, 0)$, we have

$$\begin{aligned}x &\rightarrow x' = x - vt, \\ y &\rightarrow y' = y, \\ z &\rightarrow z' = z, \\ t &\rightarrow t' = t.\end{aligned}\tag{1.36}$$

The *Galilean transformations* are the transformations connecting the coordinates of two inertial reference frames which differ only by constant relative motion and have the form (1.35).

The inverse transformation of (1.35) is

$$\begin{aligned}\mathbf{x}' &\rightarrow \mathbf{x} = \mathbf{x}' + \mathbf{v}t', \\ t' &\rightarrow t = t',\end{aligned}\tag{1.37}$$

which can be also obtained from Eq. (1.35) by replacing \mathbf{v} with $-\mathbf{v}$ and exchanging primed and unprimed coordinates.

From Eq. (1.35) we can easily infer the relation between the velocities of a particle measured in the two reference frames. If $\mathbf{w} = \dot{\mathbf{x}}$ is the velocity of the particle in the reference frame with the Cartesian coordinates (x, y, z) and $\mathbf{w}' = \dot{\mathbf{x}'}$ is the velocity of the particle in the reference frame with the Cartesian coordinates (x', y', z') , where the dot indicates the derivative with respect to time, we have

$$\mathbf{w}' = \mathbf{w} - \mathbf{v}.\tag{1.38}$$

Two inertial reference frames may also differ by a translation or a rotation. The transformation connecting two inertial reference frames which differ only by a translation is

$$\mathbf{x} \rightarrow \mathbf{x}' = \mathbf{x} + \mathbf{T},\tag{1.39}$$

where $\mathbf{T} = (T^1, T^2, T^3)$. In principle, a translation may also be applied to the temporal coordinate and we may have

$$t \rightarrow t' = t + t_0.\tag{1.40}$$

Note that the time interval measured in different reference frames is the same, namely Δt is an invariant. This is one of the key assumptions in Newtonian mechanics, where there is an absolute time valid for any observer (inertial and non-inertial).

The transformation connecting two inertial reference frames which differ only by a rotation is

$$\mathbf{x} \rightarrow \mathbf{x}' = R\mathbf{x}, \quad (1.41)$$

where R is the rotation matrix of the transformation. The rotations about the x , y , and z axes (or, equivalently, the rotations in the yz , xz , and xy planes) of the angle θ have, respectively, the following form

$$\begin{aligned} R_{yz}(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix}, \\ R_{xz}(\theta) &= \begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix}, \\ R_{xy}(\theta) &= \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \quad (1.42)$$

A general rotation can be written as a combination of elementary rotations about the x , y , and z axes, namely

$$R(\theta_{yz}, \theta_{xz}, \theta_{xy}) = R_{yz}(\theta_{yz}) \cdot R_{xz}(\theta_{xz}) \cdot R_{xy}(\theta_{xy}). \quad (1.43)$$

The Galilean transformations with the translations and the rotations form a group (see Appendix A.1), which is called the *Galilean group*. The combination of two or more transformations of coordinates is still a transformation of coordinates. The inverse element is the inverse transformation. The identity element is the trivial transformation

$$x \rightarrow x' = x, \quad y \rightarrow y' = y, \quad z \rightarrow z' = z, \quad t \rightarrow t' = t. \quad (1.44)$$

The Galilean group is the set of all possible transformations connecting different inertial frames in Newtonian mechanics. A generic transformation of the Galilean group has the form

$$\begin{aligned} \mathbf{x} \rightarrow \mathbf{x}' &= R\mathbf{x} - \mathbf{v}t + \mathbf{T}, \\ t \rightarrow t' &= t + t_0. \end{aligned} \quad (1.45)$$

Assuming that the transformations of the Galilean group are the correct transformations to move from an inertial reference frame to another inertial reference frame, the Special Principle of Relativity in Sect. 1.1 can be reformulated as follows:

Galileo's Principle of Relativity. The laws of physics are invariant under the Galilean group.

The physics of the 17 and 18th centuries was consistent with the Galileo Principle of Relativity. As shown at the end of this chapter, the situation changed in the 19th century with the study of electromagnetic phenomena. Maxwell's equations are not invariant under Galilean transformations and this was initially interpreted as an indication of the existence of a preferred reference frame. It was later realized that the problem was in the Galilean transformations, which can be used only when the relative velocity between two reference frames is much smaller than the speed of light.

1.5 Principle of Least Action

Let us consider a certain physical system. Its *action* S between the instants of time t_1 and t_2 is given by

$$S = \int_{t_1}^{t_2} L[\mathbf{q}(t), \dot{\mathbf{q}}(t), t] dt. \quad (1.46)$$

where $L[\mathbf{q}(t), \dot{\mathbf{q}}(t), t]$ is the *Lagrangian* of the system, $\mathbf{q} = (q^1, q^2, \dots, q^n)$, q^i s are the Lagrangian coordinates and define the configuration of the system, and $\dot{\mathbf{q}}$ is the derivative of \mathbf{q} with respect to t . For the moment we only say that the Lagrangian is a certain function capable of describing the dynamics of the system under consideration and we introduce the Principle of Least Action.

Principle of Least Action. The trajectory of a system between two times, say t_1 and t_2 , is the one for which its action is stationary to first order.

The Principle of Least Action is an elegant way to infer the equations of motion of a system once its action is known. We consider small changes in the configuration of the system

$$\mathbf{q}(t) \rightarrow \tilde{\mathbf{q}}(t) = \mathbf{q}(t) + \delta\mathbf{q}(t), \quad (1.47)$$

with the boundary conditions

$$\delta\mathbf{q}(t_1) = \delta\mathbf{q}(t_2) = \mathbf{0}. \quad (1.48)$$

The variation in the Lagrangian coordinates (1.47) produces a variation of the action

$$\delta S = \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{q}^i} \delta \dot{q}^i \right) dt, \quad (1.49)$$

where we have used the convention of summation over repeated indices. Since

$$\delta \dot{q}^i = \dot{\delta q}^i - \dot{q}^i = \frac{d}{dt} \delta q^i, \quad (1.50)$$

we can write

$$\frac{\partial L}{\partial \dot{q}^i} \delta \dot{q}^i = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \delta q^i \right) - \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) \delta q^i. \quad (1.51)$$

From the boundary conditions in Eq. (1.48), the first term on the right hand side in Eq. (1.51) does not give any contribution when we integrate over t . Eq. (1.49) thus becomes

$$\delta S = \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) \delta q^i dt. \quad (1.52)$$

Requiring that the action S is stationary for any small variation of the Lagrangian coordinates of the system, i.e. $\delta S = 0$ for any δq^i , we obtain the *Euler–Lagrange equations*

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0. \quad (1.53)$$

These are the equations of motion of the system.

Up to now we have not specified the Lagrangian L . However, in general there is no fundamental recipe to construct the Lagrangian of a specific system. The Lagrangian of a certain physical system is simply the one that provides the right equations of motion for that physical system. In other words, if we want to study a system, we can consider a number of Lagrangians, each of them representing a certain model for that system. We can then check (with experiments/observations) which Lagrangian can better describe the system, and thus find the best model.

The Principle of Least Action is a principle, so it cannot be proven. For the time being, all known physical systems can be treated with this formalism.

In particular classes of systems, it is straightforward to find their Lagrangians. The simplest example is a point-like particle moving in a potential V . In such a case, the Lagrangian of the system is simply given by the difference between the kinetic energy of the point-like particle T and its potential V . In three dimensions, we can write

$$L = T - V = \frac{1}{2} m \dot{\mathbf{x}}^2 - V, \quad (1.54)$$

where $\mathbf{x} = (x, y, z)$ are the Cartesian coordinates of the particle, $\dot{\mathbf{x}}$ is the particle velocity, and $\dot{\mathbf{x}}^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2$. The Euler–Lagrange equations provide the equation of motion

$$m\ddot{\mathbf{x}} = -\nabla V. \quad (1.55)$$

Equation (1.55) is the well-known Newton’s Second Law for a point-like particle in a potential V , and therefore the Lagrangian in Eq. (1.54) is the right one.

1.6 Constants of Motion

Let us now assume that the Lagrangian of a certain physical system does not explicitly depend on the time t , namely $L = L[\mathbf{q}(t), \dot{\mathbf{q}}(t)]$. In such a case

$$\frac{\partial L}{\partial t} = 0, \quad (1.56)$$

and therefore

$$\frac{dL}{dt} = \frac{\partial L}{\partial q^i} \dot{q}^i + \frac{\partial L}{\partial \dot{q}^i} \ddot{q}^i + \frac{\partial L}{\partial t} = \frac{\partial L}{\partial q^i} \dot{q}^i + \frac{\partial L}{\partial \dot{q}^i} \ddot{q}^i, \quad (1.57)$$

which we can rewrite as

$$\frac{\partial L}{\partial q^i} \dot{q}^i + \frac{\partial L}{\partial \dot{q}^i} \ddot{q}^i - \frac{dL}{dt} = 0. \quad (1.58)$$

From the Euler–Lagrange equations

$$\frac{\partial L}{\partial q^i} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i}, \quad (1.59)$$

and Eq. (1.58) can be rewritten as

$$\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) \dot{q}^i + \frac{\partial L}{\partial \dot{q}^i} \ddot{q}^i - \frac{dL}{dt} = 0, \quad (1.60)$$

and also as

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \dot{q}^i - L \right) = 0. \quad (1.61)$$

The expression in brackets in Eq. (1.61) is a constant of motion

$$E = \frac{\partial L}{\partial \dot{q}^i} \dot{q}^i - L . \quad (1.62)$$

In the case of a point-like particle moving in the potential V , we have

$$\frac{\partial L}{\partial \dot{q}^i} \dot{q}^i = m \dot{\mathbf{x}}^2 = 2T , \quad (1.63)$$

and

$$E = T + V . \quad (1.64)$$

E is the energy of the point-like particle.

Let us now consider the case in which the Lagrangian of a system does not depend on a certain Lagrangian coordinate, say q^i . Since $\partial L / \partial q^i = 0$, from the Euler–Lagrange equations it follows that

$$\frac{d}{dt} p_i = 0 , \quad (1.65)$$

where p_i is the *conjugate momentum* defined as

$$p_i = \frac{\partial L}{\partial \dot{q}^i} . \quad (1.66)$$

p_i is a constant of motion of the system.

The simplest example is that of a free point-like particle. Since $V = 0$, the Lagrangian of the system is just the kinetic energy of the particle. In three dimensions and with Cartesian coordinates, we have

$$L = \frac{1}{2} m (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) . \quad (1.67)$$

The constants of motion are the three components of the momentum

$$p_x = m\dot{x} , \quad p_y = m\dot{y} , \quad p_z = m\dot{z} , \quad (1.68)$$

as well as the energy $E = T$.

1.7 Geodesic Equations

In Newtonian mechanics, the Lagrangian of a free point-like particle is simply the particle kinetic energy and is given in Eq. (1.67). The equations of motion can be obtained by minimizing the action

$$S = \frac{1}{2}m \int_{\Gamma} (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) dt, \quad (1.69)$$

where Γ is the particle trajectory. The Euler–Lagrange equations are $\ddot{x} = \ddot{y} = \ddot{z} = 0$ and the solution is a constant velocity along a straight line.

If we consider a non-Cartesian coordinate system, for example spherical coordinates (r, θ, ϕ) , the action can be written as

$$S = \frac{1}{2}m \int_{\Gamma} g_{ij} \dot{x}^i \dot{x}^j dt, \quad (1.70)$$

where g_{ij} is the metric tensor introduced in Sect. 1.2. Note that the line element $dl^2 = g_{ij} dx^i dx^j$ is an invariant, but the square of the velocity $v^2 = g_{ij} \dot{x}^i \dot{x}^j$ is an invariant only if we do not consider reference frames with non-vanishing relative motion. This is simply because in dl^2 we are considering the infinitesimal distance between two specific points of the space, say \mathbf{x}_A and \mathbf{x}_B , which exist independently of the coordinate system and have different coordinates in different reference frames. In v^2 , $\dot{x}^i = dx^i/dt$ where dx^i is the change of the values of coordinates in the time dt in a certain reference frame, but the points of the space are different in the two reference frames with non-vanishing relative velocity.

Now the Euler–Lagrange equations are

$$\begin{aligned} \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^k} - \frac{\partial L}{\partial x^k} &= 0, \\ \frac{d}{dt} \left(g_{ij} \delta_k^i \dot{x}^j + g_{ij} \delta_k^j \dot{x}^i \right) - \frac{\partial g_{ij}}{\partial x^k} \dot{x}^i \dot{x}^j &= 0, \\ \frac{d}{dt} \left(2g_{ik} \dot{x}^i \right) - \frac{\partial g_{ij}}{\partial x^k} \dot{x}^i \dot{x}^j &= 0, \\ 2g_{ik} \ddot{x}^i + 2 \frac{\partial g_{ik}}{\partial x^j} \dot{x}^i \dot{x}^j - \frac{\partial g_{ij}}{\partial x^k} \dot{x}^i \dot{x}^j &= 0, \\ 2g_{ik} \ddot{x}^i + \frac{\partial g_{ik}}{\partial x^j} \dot{x}^i \dot{x}^j + \frac{\partial g_{jk}}{\partial x^i} \dot{x}^i \dot{x}^j - \frac{\partial g_{ij}}{\partial x^k} \dot{x}^i \dot{x}^j &= 0. \end{aligned} \quad (1.71)$$

We multiply the last equation in (1.71) by g^{lk} (remember that $g^{ij} g_{jk} = \delta_k^i$)

$$\delta_k^i \ddot{x}^i + \frac{1}{2} g^{lk} \left(\frac{\partial g_{ik}}{\partial x^j} \dot{x}^i \dot{x}^j + \frac{\partial g_{jk}}{\partial x^i} \dot{x}^i \dot{x}^j - \frac{\partial g_{ij}}{\partial x^k} \dot{x}^i \dot{x}^j \right) = 0. \quad (1.72)$$

The final equation can be written as

$$\ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k = 0, \quad (1.73)$$

which is called the *geodesic equations*. Γ_{jk}^i s are the *Christoffel symbols*

$$\Gamma_{jk}^i = \frac{1}{2} g^{il} \left(\frac{\partial g_{lk}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l} \right). \quad (1.74)$$

As we will see in Sect. 5.2.1, the Christoffel symbols are not the components of a tensor.

Note that the geodesic equations can be obtained even if we apply the Least Action Principle to the length of the trajectory of the particle

$$\ell = \int_{\Gamma} dl = \int_{\Gamma} \sqrt{g_{ij} \dot{x}^i \dot{x}^j} dt. \quad (1.75)$$

Indeed, modulo a constant, the new Lagrangian is $L' = \sqrt{L}$ and the Euler–Lagrange equations for L' are

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2\sqrt{L}} \frac{\partial L}{\partial \dot{x}^k} \right) - \frac{1}{2\sqrt{L}} \frac{\partial L}{\partial x^k} &= 0, \\ -\frac{1}{4L^{3/2}} \frac{dL}{dt} \frac{\partial L}{\partial \dot{x}^k} + \frac{1}{2\sqrt{L}} \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^k} - \frac{1}{2\sqrt{L}} \frac{\partial L}{\partial x^k} &= 0. \end{aligned} \quad (1.76)$$

Since L is (modulo a constant) the particle kinetic energy and is conserved for a free particle, $dL/dt = 0$, and we recover the Euler–Lagrange equations for L .⁴

If we have a metric tensor g_{ij} , we can compute the Christoffel symbols from their definition (1.74). Nevertheless, it is usually faster to write the Euler–Lagrange equations for the motion of a free point-like particle and identify the non-vanishing Christoffel symbols by comparing the result with the geodesic equations. As an example, let us consider the metric tensor in (1.15). The Lagrangian of a free point-like particle is

$$L = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \sin^2 \theta \dot{\phi}^2). \quad (1.77)$$

The Euler–Lagrange equation for the r coordinate is

$$\ddot{r} - r \dot{\theta}^2 - r \sin^2 \theta \dot{\phi}^2 = 0. \quad (1.78)$$

For the θ coordinate, we have

$$\begin{aligned} \frac{d}{dt} (r^2 \dot{\theta}) - r^2 \sin \theta \cos \theta \dot{\phi}^2 &= 0, \\ 2r \dot{r} \dot{\theta} + r^2 \ddot{\theta} - r^2 \sin \theta \cos \theta \dot{\phi}^2 &= 0, \end{aligned} \quad (1.79)$$

⁴If we think of t as the parameter that parametrizes the curve (rather than the time coordinate), we can always choose t such that L is constant and thus recover the geodesic equations. Of course, the choice of the parametrization does not affect the solution of the equations. It only simplifies the equations to solve.

which can be rewritten as

$$\ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} - \sin\theta \cos\theta \dot{\phi}^2 = 0. \quad (1.80)$$

Lastly, for the ϕ coordinate we have

$$\begin{aligned} \frac{d}{dt}(r^2 \sin^2 \theta \dot{\phi}) &= 0, \\ 2r\dot{r} \sin^2 \theta \dot{\phi} + 2r^2 \sin\theta \cos\theta \dot{\theta} \dot{\phi} + r^2 \sin^2 \theta \ddot{\phi} &= 0, \end{aligned} \quad (1.81)$$

which we rewrite as

$$\ddot{\phi} + \frac{2}{r}\dot{r}\dot{\phi} + 2\cot\theta \dot{\theta} \dot{\phi} = 0. \quad (1.82)$$

If we compare Eqs. (1.78), (1.80), and (1.82) with the geodesic equations (1.73), we see that the non-vanishing Christoffel symbols are

$$\begin{aligned} \Gamma_{\theta\theta}^r &= -r, & \Gamma_{\phi\phi}^r &= -r \sin^2 \theta, \\ \Gamma_{r\theta}^\theta &= \Gamma_{\theta r}^\theta = \frac{1}{r}, & \Gamma_{\phi\phi}^\theta &= -\sin\theta \cos\theta, \\ \Gamma_{r\phi}^\phi &= \Gamma_{\phi r}^\phi = \frac{1}{r}, & \Gamma_{\theta\phi}^\phi &= \Gamma_{\phi\theta}^\phi = \cot\theta. \end{aligned} \quad (1.83)$$

1.8 Newton's Gravity

Let us consider a point-like test-particle of mass m moving in the gravitational field of a point-like massive body of mass M ($M \gg m$). The Lagrangian of the point-like test-particle is

$$L = \frac{1}{2}m\dot{\mathbf{x}}^2 - m\Phi = \frac{1}{2}m\dot{\mathbf{x}}^2 + \frac{G_N M m}{r}, \quad (1.84)$$

where $\dot{\mathbf{x}}$ is the velocity of the particle, Φ is the gravitational potential, and r is the distance of the particle from the massive body. In Cartesian coordinates, we have

$$\dot{\mathbf{x}}^2 = \delta_{ij}\dot{x}^i \dot{x}^j, \quad r = \sqrt{x^2 + y^2 + z^2}. \quad (1.85)$$

Cartesian coordinates are not convenient for such a system with spherical symmetry. While the physics is independent of the choice of coordinates, it is easier to study the system in spherical coordinates. In spherical coordinates we have

$$L = \frac{1}{2}m (\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2 \theta \dot{\phi}^2) + \frac{G_N M m}{r}. \quad (1.86)$$

The Euler–Lagrange equation for the θ coordinate is

$$\frac{d}{dt} (mr^2\dot{\theta}) - mr^2 \sin \theta \cos \theta \dot{\phi}^2 = 0. \quad (1.87)$$

If the motion of the particle is initially in the equatorial plane, namely $\theta(t_0) = \pi/2$ and $\dot{\theta}(t_0) = 0$, where t_0 is some initial time, it remains in the equatorial plane. Without loss of generality, we can thus study the case of a particle moving in the equatorial plane (if this were not the case, we can always perform a proper rotation of the coordinate system to meet such a condition). The Lagrangian of the particle can thus be simplified to the form

$$L = \frac{1}{2}m (\dot{r}^2 + r^2\dot{\phi}^2) + \frac{G_N M m}{r}. \quad (1.88)$$

Since the Lagrangian does not explicitly depend on the coordinate ϕ and the time t , we have two constants of motion, which are, respectively, the axial component of the angular momentum L_z and the energy E . Following the approach of Sect. 1.6, we have⁵

$$\frac{d}{dt} (mr^2\dot{\phi}) = 0 \quad \Rightarrow \quad mr^2\dot{\phi} = \text{const.} = L_z, \quad (1.89)$$

and

$$\begin{aligned} E &= \frac{1}{2}m\dot{\mathbf{x}}^2 - \frac{G_N M m}{r} \\ &= \frac{1}{2}m (\dot{r}^2 + r^2\dot{\phi}^2) - \frac{G_N M m}{r} \\ &= \frac{1}{2}m\dot{r}^2 + \frac{L_z^2}{2mr^2} - \frac{G_N M m}{r}. \end{aligned} \quad (1.90)$$

If we define $\tilde{E} \equiv E/m$ and $\tilde{L}_z \equiv L_z/m$, we can write the following equation of motion

$$\frac{1}{2}\dot{r}^2 = \tilde{E} - V_{\text{eff}}, \quad (1.91)$$

where V_{eff} is the effective potential

$$V_{\text{eff}} = -\frac{G_N M}{r} + \frac{\tilde{L}_z^2}{2r^2}. \quad (1.92)$$

⁵We use the notation L_z because this is the axial component of the angular momentum and we do not want to call it L because it may generate confusion with the Lagrangian.

Fig. 1.2 Plot of the effective potential V_{eff} in Eq. (1.92) for $G_N M = 1$ and $\tilde{L}_z = 3.9$

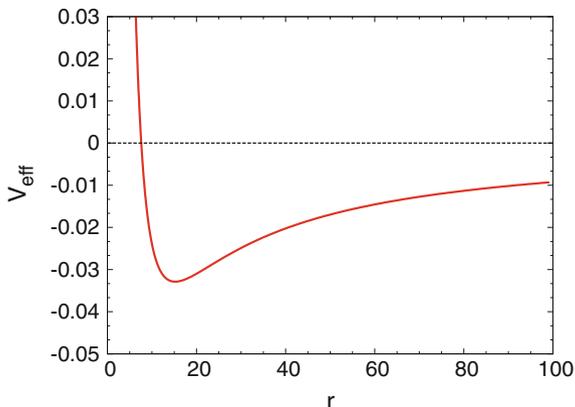


Figure 1.2 shows V_{eff} as a function of the radial coordinate r . The first term on the right hand side in (1.92) is the (attractive) gravitational potential of a point-like body of mass M and is dominant when the particle is at large radii. The second term is the (repulsive) centrifugal potential and is dominant when the particle is at small radii. In Newton's gravity, the centrifugal potential prevents a point-like particle with non-vanishing angular momentum from falling onto a point-like massive body. As shown in Sect. 8.4, this is not true in Einstein's gravity.

1.9 Kepler's Laws

Kepler's Laws were empirically discovered by Johannes Kepler at the beginning of the 17th century by studying astronomical data of the planets in the Solar System. It was later shown by Isaac Newton that these laws are a direct consequence of Newtonian mechanics and Newton's Law of Universal Gravitation. Kepler's Laws read as follows:

Kepler's First Law. The orbit of a planet is an ellipse with the Sun at one of the two foci.

Kepler's Second Law. The line connecting a planet to the Sun sweeps out equal areas in equal times.

Kepler's Third Law. The square of the period of any planet is proportional to the cube of the semi-major axis of its orbit.

The Lagrangian describing the motion of a planet of mass m orbiting the Sun with mass M ($m \ll M$) is given in Eq. (1.88). Kepler's Second Law is the direct consequence of the fact that the angular momentum of the planet is a constant of

motion. The area swept out by the line connecting the planet to the Sun is indeed given by

$$A(\phi_1, \phi_2) = \frac{1}{2} \int_{\phi_1}^{\phi_2} r^2 d\phi = \frac{1}{2} \int_{t_1}^{t_2} r^2 \dot{\phi} dt = \frac{L_z}{2m} \int_{t_1}^{t_2} dt = \frac{L_z}{2m} (t_2 - t_1), \quad (1.93)$$

where ϕ_1 and ϕ_2 are the values of the ϕ coordinate at the time t_1 and t_2 , respectively.

Kepler's First Law can be derived as follows. We rewrite Eq. (1.90) by replacing \dot{r} with

$$\dot{r} = \frac{dr}{d\phi} \frac{d\phi}{dt} = \frac{L_z}{mr^2} \frac{dr}{d\phi}, \quad (1.94)$$

and we find

$$E = \frac{L_z^2}{2mr^4} \left(\frac{dr}{d\phi} \right)^2 + \frac{L_z^2}{2mr^2} - \frac{G_N M m}{r}. \quad (1.95)$$

We define

$$r = \frac{1}{u}, \quad u' = \frac{du}{d\phi}, \quad (1.96)$$

and Eq. (1.95) becomes

$$E = \frac{L_z^2}{2m} u'^2 + \frac{L_z^2}{2m} u^2 - G_N M m u, \quad (1.97)$$

and then

$$u'^2 - \frac{2G_N M m^2}{L_z^2} u + u^2 = \frac{2mE}{L_z^2}. \quad (1.98)$$

We derive Eq. (1.98) with respect to ϕ and we find

$$2u' \left(u'' - \frac{G_N M m^2}{L_z^2} + u \right) = 0. \quad (1.99)$$

We have thus two equations:

$$u' = 0, \quad (1.100)$$

$$u'' - \frac{G_N M m^2}{L_z^2} + u = 0. \quad (1.101)$$

The solutions of the differential equations in (1.100) and (1.101) are, respectively, the equations of a circle and of a conic

$$\frac{1}{r} = u = \text{constant} , \quad (1.102)$$

$$\frac{1}{r} = u = \frac{G_N M m^2}{L_z^2} + A \cos \phi , \quad (1.103)$$

where A is a constant. Eq. (1.103) describes an ellipse if $0 < A < G_N M m^2 / L_z^2$ [see Eqs. (D.6) and (D.7) in Appendix D], a parabola if $A = G_N M m^2 / L_z^2$, and a hyperbola if $A > G_N M m^2 / L_z^2$.

From Eq. (1.93), we can write the area of the ellipse described by the orbit of the planet as

$$\pi a^2 \sqrt{1 - e^2} = \frac{1}{2} \int_0^{2\pi} r^2 d\phi = \frac{L_z}{2m} \int_0^T dt = \frac{L_z T}{2m} , \quad (1.104)$$

where a is the ellipse semi-major axis, e is the ellipse eccentricity, and T is the orbital period of the planet. From Eqs. (D.6) and (D.7) in Appendix D

$$\frac{G_N M m^2}{L_z^2} = \frac{1}{a(1 - e^2)} \Rightarrow L_z^2 = a(1 - e^2) G_N M m^2 . \quad (1.105)$$

Combining Eq. (1.104) with Eq. (1.105), we find Kepler's Third Law

$$\frac{a^3}{T^2} = \frac{G_N M}{4\pi^2} = \text{constant} \Rightarrow a^3 \propto T^2 . \quad (1.106)$$

1.10 Maxwell's Equations

The physics of the 17th and 18th centuries was in perfect agreement with the Principle of Relativity, according to which there is no preferred reference frame and all inertial reference frames are equivalent and related to each other by the transformations of the Galilean group. However, the situation changed in the 19th century with the study of electromagnetic phenomena. Maxwell's equations in vacuum are (in Gaussian units)

$$\nabla \cdot \mathbf{E} = 0 , \quad (1.107)$$

$$\nabla \cdot \mathbf{B} = 0 , \quad (1.108)$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} , \quad (1.109)$$

$$\nabla \times \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} , \quad (1.110)$$

and they look to be inconsistent with the Principle of Relativity. Here c is the speed of light. In the context of Galileo's relativity, this is a physical quantity that changes according to Eq. (1.38) if we change inertial reference frame, which suggests that Maxwell's equations must hold in some special inertial reference frame and not in the others. Moreover, Maxwell's equations are not invariant under Galilean transformations. There are thus two possibilities:

1. The Principle of Relativity does not hold for electromagnetic phenomena, which require a preferred reference frame.
2. The laws of physics are the same in all inertial reference frames, but the Galilean transformations are wrong.

It is easy to see from Maxwell's equations in vacuum that electromagnetic phenomena have wave properties. For a generic vector \mathbf{V} , we have the following identity (see Appendix B.3)

$$\nabla \times (\nabla \times \mathbf{V}) = \nabla (\nabla \cdot \mathbf{V}) - \nabla^2 \mathbf{V}. \quad (1.111)$$

From Maxwell's third equation (1.109), we can write

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left(-\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \right). \quad (1.112)$$

We rewrite the left hand side of this equation by employing Eq. (1.111) and Maxwell's first equation (1.107)

$$\nabla^2 \mathbf{E} = \frac{1}{c} \frac{\partial}{\partial t} (\nabla \times \mathbf{B}), \quad (1.113)$$

where the right hand side has been rewritten by exploiting the fact that $\partial/\partial t$ and $\nabla \times$ commute. We then use Eq. (1.110) and we obtain

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{E} = \mathbf{0}. \quad (1.114)$$

With a similar procedure, we can also write

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{B} = \mathbf{0}. \quad (1.115)$$

Equations (1.114) and (1.115) are wave equations for the fields \mathbf{E} and \mathbf{B} . In Newtonian mechanics, waves arise from the displacement of a portion of an elastic medium from its natural position, which then starts oscillating about its equilibrium position. A mechanical wave is a perturbation of the elastic medium and propagates with velocity

$$v = \sqrt{\frac{T}{\rho}}, \quad (1.116)$$

where T and ρ are, respectively, the tension and the density of the elastic medium.

At the end of the 19th century, it was natural to postulate the existence of a medium for the propagation of electromagnetic phenomena. Such a medium was called the *aether* and the speed of the electromagnetic phenomena was c . The aether had to be the preferred reference frame for the description of electromagnetic phenomena. At this point, it was necessary to find direct or indirect evidence for the existence of the aether.

1.11 Michelson–Morley Experiment

The Michelson-Morley experiment is an interferometer used to measure the difference of the speed of light between two orthogonal directions as the result of the possible motion of the set-up with respect to the aether. The set-up is sketched in Fig. 1.3. The source S emits a beam of light, which is separated into two beams by a beamsplitter B . One of the light beams goes to branch 1, is reflected by mirror M_1 , and returns to beamsplitter B . The second light beam goes to branch 2, is reflected by mirror M_2 , and returns to beamsplitter B . The two beams then go to a detector D , where we can observe the interference pattern. As shown in Fig. 1.3, we indicate with d_1 and d_2 the lengths of branch 1 and branch 2, respectively.

Let us assume that the aether is moving with the speed v parallel to branch 1 (configuration A). According to the Galilean transformations, the travel time of the light beam in branch 1, namely the time that light takes to go from beamsplitter B to mirror M_1 and return to beamsplitter B , would be

$$\Delta t_1^A = \frac{d_1}{c-v} + \frac{d_1}{c+v} = \frac{2d_1c}{c^2-v^2} = \frac{2d_1}{c} \left[1 + \frac{v^2}{c^2} + O\left(\frac{v^4}{c^4}\right) \right], \quad (1.117)$$

where c is the speed of light in the reference frame of the aether. In the reference frame of the aether, the travel distance of the light beam in branch 2 would be

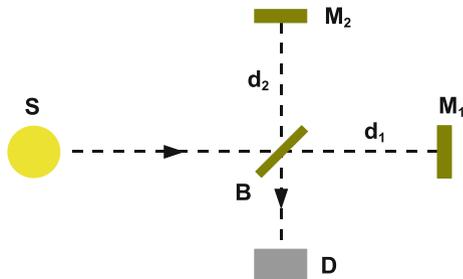
$$D = \left[4d_2^2 + v^2 (\Delta t_2^A)^2 \right]^{1/2}, \quad (1.118)$$

where Δt_2^A is the time that the light beam takes to go from B to M_2 and return to B . Since the speed of light in the reference frame of the aether is c , Δt_2^A is given by

$$\Delta t_2^A = \frac{D}{c}. \quad (1.119)$$

If we combine Eq. (1.118) with Eq. (1.119), we get Δt_2^A

Fig. 1.3 Sketch of the Michelson-Morley experiment. S is the source, B is the beamsplitter, M_1 and M_2 are the mirrors in branches 1 and 2, respectively, and D is the detector where we can observe the interference pattern. d_1 and d_2 are the lengths of branches 1 and 2, respectively



$$\Delta t_2^A = \frac{2d_2}{(c^2 - v^2)^{1/2}} = \frac{2d_2}{c} \left[1 + \frac{1}{2} \frac{v^2}{c^2} + O\left(\frac{v^4}{c^4}\right) \right]. \quad (1.120)$$

The time difference of the two signals is

$$\delta t^A = \Delta t_1^A - \Delta t_2^A = 2 \frac{d_1 - d_2}{c} + (2d_1 - d_2) \frac{v^2}{c^3} + O\left(\frac{v^4}{c^5}\right). \quad (1.121)$$

Let us now rotate the set-up by 90° (configuration B) and evaluate the time difference in the new configuration. Now the aether should move with the speed v parallel to branch 2. We find

$$\delta t^B = \Delta t_1^B - \Delta t_2^B = 2 \frac{d_1 - d_2}{c} + (d_1 - 2d_2) \frac{v^2}{c^3} + O\left(\frac{v^4}{c^5}\right). \quad (1.122)$$

The hypothesis of the existence of the aether together with the Galilean transformations predict the following fringe shift of the interference pattern between configurations A and B

$$\delta n = \frac{\delta T}{T} = \frac{c}{\lambda} (\delta t^A - \delta t^B) = \frac{d_1 + d_2}{\lambda} \frac{v^2}{c^2} + O\left(\frac{v^4}{c^4}\right), \quad (1.123)$$

where $T = \lambda/c$ is the period of the radiation and λ is its wavelength.

The first attempt to measure the speed of the aether with respect to Earth was done in 1881 by Albert Michelson, but the result was not conclusive. The experiment was repeated in 1887 by Albert Michelson and Edward Morley. In the 1887 experiment, $d_1 = d_2 = 11$ m and $\lambda \approx 600$ nm. If we assume that the rest frame of the aether is that of the Solar System and we consider that the orbital velocity of Earth is about 30 km/s, we should expect $\delta n = 0.4$. If we consider the velocity of the Solar System around the galactic center, which is about 220 km/s, we should expect $\delta n = 3$. The resolution of the 1887 experiment was $\delta n = 0.01$ and no difference in the interference pattern was observed between the two orientations of the set-up. The experiment was repeated by other physicists with the same negative result. Strictly speaking, these experiments do not rule out the hypothesis of the existence of the aether if we postulate that the space

is not isotropic with respect to length measurements, and such a possibility was the first solution explored by physicists to save the aether and explain the experimental results. However, similar attempts eventually failed.

1.12 Towards the Theory of Special Relativity

In 1887 Woldemar Voigt wrote the following coordinate transformation relating an inertial reference frame with Cartesian coordinates (x, y, z) to another inertial reference frame with Cartesian coordinates (x', y', z') and moving in the positive x -direction at constant speed v relative to the former reference frame

$$\begin{aligned} x &\rightarrow x' = x - vt, \\ y &\rightarrow y' = y \left(1 - \frac{v^2}{c^2}\right)^{1/2}, \\ z &\rightarrow z' = z \left(1 - \frac{v^2}{c^2}\right)^{1/2}, \\ t &\rightarrow t' = t - \frac{vx}{c^2}. \end{aligned} \tag{1.124}$$

Such a coordinate transformation is a modification of the Galilean transformation in Eq. (1.36). Voigt showed that some electrodynamics equations were invariant under the transformation in (1.124). For instance, the electromagnetic wave Eqs. (1.114) and (1.115) become

$$\begin{aligned} \left(1 - \frac{v^2}{c^2}\right) \left(\nabla'^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t'^2}\right) \mathbf{E} &= \mathbf{0}, \\ \left(1 - \frac{v^2}{c^2}\right) \left(\nabla'^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t'^2}\right) \mathbf{B} &= \mathbf{0}. \end{aligned} \tag{1.125}$$

However, it seems he did not attribute any particular physical meaning to this coordinate transformation.

In 1887, Michelson and Morley announced the results of their experiment. Unlike that in 1881, the result was convincing. However, the hypothesis of the existence of the aether was not immediately abandoned. In 1889, George FitzGerald and, independently, in 1892, Hendrik Lorentz showed that the Michelson–Morley experiment could be explained postulating the contraction of the length L of an object moving parallel to the aether (*FitzGerald–Lorentz contraction*)

$$L \rightarrow L \left(1 - \frac{v^2}{c^2}\right)^{1/2}. \tag{1.126}$$

In 1897, Joseph Larmor extended Lorentz's work, writing the coordinate transformation of special relativity between two inertial reference frames differing by a constant relative speed, and showed that the FitzGerald-Lorentz contraction was a consequence of this coordinate transformation. In 1905, Henri Poincaré gave these coordinate transformations their modern form and called them the Lorentz transformations. Poincaré also showed that these transformations together with the rotations form a group, which was called the Lorentz group. In the same year, Albert Einstein showed that these transformations can be derived assuming the Principle of Relativity and the constancy of the speed of light.

Problems

1.1 Verify Eq. (1.14).

1.2 The transformation between spherical coordinates (r, θ, ϕ) and cylindrical coordinates (ρ, z, ϕ') is

$$\rho = r \sin \theta, \quad z = r \cos \theta, \quad \phi' = \phi, \quad (1.127)$$

with inverse

$$r = \sqrt{\rho^2 + z^2}, \quad \theta = \arctan\left(\frac{\rho}{z}\right), \quad \phi = \phi'. \quad (1.128)$$

Write the metric tensor g_{ij} and then the line element dl in cylindrical coordinates.

1.3 Consider the transformation $x^i \rightarrow x'^i$ described by the Galilean transformation in Eq. (1.36). Show that the expression of the Euclidean metric δ_{ij} does not change.

1.4 Consider the transformation $x^i \rightarrow x'^i$ described by the rotation in the xy plane R_{xy} in Eq. (1.42). Show that the expression of the Euclidean metric δ_{ij} does not change.

1.5 The Lagrangian of a free point-like particle in spherical coordinates is in Eq. (1.77) and the transformations between spherical coordinates (r, θ, ϕ) and cylindrical coordinates (ρ, z, ϕ') are given by Eqs. (1.127) and (1.128). Write the Lagrangian in cylindrical coordinates and then the corresponding Euler–Lagrange equations.

1.6 From the Euler–Lagrange equations obtained in the previous exercise, write the Christoffel symbols in cylindrical coordinates.

1.7 The Lagrangian of a free point-like particle of mass m moving on a spherical surface of radius R is

$$L = \frac{1}{2}mR^2 (\dot{\theta}^2 + \sin^2 \theta \dot{\phi}^2) . \quad (1.129)$$

Note that here the Lagrangian coordinates are (θ, ϕ) , while R is a constant. Write the Euler–Lagrange equations for the Lagrangian in (1.129).

1.8 Let us consider the following Lagrangian

$$L = \frac{1}{2}m (\dot{x}^2 + \dot{y}^2) - \frac{1}{2}k (x^2 + y^2) , \quad (1.130)$$

where x and y are the Lagrangian coordinates. It is the Lagrangian of a particle moving in a 2-dimensional space and subject to the potential $V = k(x^2 + y^2)/2$. Find the constant(s) of motion and then write the corresponding Euler–Lagrange equations.

1.9 Show that the Maxwell equations are not invariant under Galilean transformations.