1

THE GRAVITATIONAL FIELD

In this chapter the basic and preliminary properties of the gravitational phenomena are given. These are the prerelativistic properties which lay the foundations of the theory of general relativity. The discussion starts with the Newtonian theory of gravitation, along with other related topics, such as Newton's laws of motion. It then proceeds to the concepts of gravitational and inertial forces and their mutual relationship. This is followed by a discussion of the equality of the gravitational mass to the inertial mass, along with the experimental verification of this important fact. The experiment, known as the Eötvös experiment, is subsequently examined in detail. The chapter is concluded by discussing the principle of equivalence and the principle of general covariance. These two principles were the basis for the physical foundations in the original formulation of the theory of general relativity by Einstein.

1.1 NEWTONIAN GRAVITATION

The Galilean Group

In the classical mechanics of Newton we assume that the laws of motion do not depend on the choice of a particular fixed system of coordinates with respect to which the distances, velocities, accelerations, forces, and so on, are being measured. Furthermore, we assume that the laws of motion do not change their forms by transferring from one system of coordinates into another. These systems of coordinates are all assumed to have uniform, rectilinear, translational motions with respect to each other. They are called *inertial systems of coordinates.*

Thus inertial systems of coordinates differ from one another by orthogonal rotations, accompanied by translations of the origins of the systems, and by motion in uniform velocities. We can further add the translation of the time parameter, namely, the possibility of choosing the origin of time, $t = 0$, at will. We may count the number of parameters, or the number of degrees of freedom, which each coordinate system has with respect to any other one. Thus we have four parameters which account for the translations of the three spatial coordinates and time, three parameters describing the orthogonal rotations of the spatial coordinates, and finally three more parameters accounting for the rectilinear motions of the spatial coordinates. Newtonian laws of classical mechanics are therefore invariant under all of these ten-parameter transformations of inertial systems of coordinates.

A transformation of inertial coordinates having ten parameters, as described above, is called a *Galilean transformation*. Newton's classical laws of mechanics are invariant under the ten-parameter Galilean transformations. We say in this case that we have a *Galilean invariance*. The aggregate of all Galilean transformations forms a group. This group is called the *Galilean group* and has ten parameters.

If we choose two inertial coordinate systems so that their corresponding axes are parallel and coincide at $t = 0$, and if $\mathbf{v}$ is the velocity of one inertial coordinate system with respect to the other, the Galilean transformation can then be reduced into a simple transformation as follows:

$$x' = x + v_x t, \qquad y' = y + v_y t, \qquad z' = z + v_z t. \tag{1.1.1}$$

Here v_x, v_y, and v_z are the components of velocity $\mathbf{v}$ along the x axis, y axis, and z axis, respectively.

Newtonian Mechanics

The Newtonian laws of mechanics are based on three fundamental laws. These laws can be stated as follows:

1. A particle acted upon by no force will assume a rectilinear motion with a constant velocity.
2. A particle acted upon by a force $\mathbf{f}$ will move with an acceleration $\mathbf{a}$ which is proportional to the force. We can then write the relation between the force and the acceleration in the form of Newton's familiar law of motion:

   $$\mathbf{f} = m\mathbf{a}, \tag{1.1.2}$$

   where m is the mass of the particle.
3. For each action there is a reaction which is equal to the action, but is directed in the opposite direction of the action.

We conclude these brief remarks on Newtonian mechanics by mentioning the concept of *action-at-a-distance* which the Newtonian theory assumes. Roughly speaking, action-at-a-distance means that interactions between particles take place instantly. This is in contrast to modern physics concepts where we assume that interactions are mediated through intermediate particles, thus leading to the concept of fields.

Newton's Theory of Gravitation

Newton's theory of gravitation is actually a three-dimensional field theory. The gravitational field is assumed to be described by a scalar field $\phi(x, y, z)$, which is a function of the spatial coordinates. The function $\phi(x, y, z)$ satisfies a second-order partial differential equation of the form

$$\nabla^2\phi(x, y, z) = 4\pi G\rho(x, y, z). \tag{1.1.3}$$

Such an equation is called the *Poisson equation*. Here G is Newton's gravitational constant, whose value is equal to $6.67 \times 10^{-8}\ \mathrm{cm}^3 \cdot \mathrm{g}^{-1} \cdot \mathrm{s}^{-2}$ in CGS units, and $\rho(x, y, z)$ is the mass density of the matter in space producing the gravitational field. The differential operator ∇^2 is given by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1.1.4}$$

and is called the *Laplacian operator*.

A solution of the Poisson equation gives the potential $\phi(x, y, z)$ in terms of the mass distribution $\rho(x, y, z)$ in space. At points where there is no matter, that is, at points of space where $\rho(x, y, z) = 0$, we can solve the equation

$$\nabla^2\phi(x, y, z) = 0. \tag{1.1.5}$$

The latter equation is called the *Laplace equation*. Its solution then describes the Newtonian potential at points of space where the mass density ρ vanishes.

The Newtonian potential ϕ creates a force field that acts on particles. This gravitational field of forces is proportional to the negative of the gradient of potential ϕ. Hence the force acting on a particle with mass m, located in a Newtonian potential ϕ, is given by

$$\mathbf{F} = -m\nabla\phi, \tag{1.1.6}$$

where ∇ is the three-dimensional gradient operator,

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right). \tag{1.1.7}$$

For instance, if the potential ϕ is produced by a single mass M, then the solution of the Poisson equation yields

$$\phi = -\frac{GM}{r}, \tag{1.1.8}$$

and the force acting on another particle with mass m will be

$$\mathbf{F} = GmM\nabla\frac{1}{r} = -\frac{GmM}{r^2}. \tag{1.1.9}$$

Equation (1.1.9) is the familiar inverse-square *law of interaction* of Newton.

The masses m and M appearing in Eq. (1.1.9) are the *gravitational masses*, since they give the gravitational attraction force between the two particles. The mass appearing in Newton's second law, Eq. (1.1.2), on the other hand, is the *inertial mass* of the particle. In Newtonian physics these two concepts are identified. In the sequel we will find that this identification is valid in general relativity theory, too.

Finally the *potential energy* for an arbitrary mass distribution in the Newtonian theory can be found. The potential energy of a particle in a gravitational field is equal to its mass times the potential of the field. Hence we obtain for the potential energy of a general system, with mass density ρ, the following expression:

$$U = \frac{1}{2}\int \rho\phi\, d^3x. \tag{1.1.10}$$

In the next section we discuss more thoroughly the basic properties of the gravitational field. This is done from a more general point of view and not necessarily that of Newtonian physics.

PROBLEMS

1.1.1 Find the Newtonian potential produced by a system of masses at distances that are large compared to the dimensions of the system.

Solution: The Newtonian potential is the solution of the Poisson equation

$$\nabla^2\phi(x) = 4\pi G\rho(x), \tag{1}$$

where $\rho(x)$ is the mass density of the system, and G is Newton's gravitational constant. In Eq. (1) the variable x denotes the three spatial coordinates x, y, z.

The solution of Eq. (1) is given by

$$\phi(x) = -G\int \frac{\rho(x')}{|\mathbf{r} - \mathbf{r}'|} d^3x', \tag{2}$$

where $\mathbf{r} = (x^1, x^2, x^3)$ is the radius vector of the point where the potential is being calculated, and $\mathbf{r}' = (x'^1, x'^2, x'^3)$ is the radius vector of an arbitrary point at the mass distribution of the matter.

The potential $\phi(x)$ can be expanded in powers of $1/r$, thus getting

$$\phi = -G\left[\frac{m}{r} + \frac{1}{6}D_{ij}\frac{\partial^2}{\partial x^i\,\partial x^j}\left(\frac{1}{r}\right) + \cdots\right], \tag{3}$$

where

$$m = \int \rho\, d^3x \tag{4}$$

is the total mass of the system. The missing $1/r^2$ term, corresponding to the dipole moment of the system of masses, vanishes identically in virtue of choosing the origin of the coordinates at the center of masses. The quantity

$$D_{ij} = \int \rho(3x^i x^j - r^2\delta^{ij})\, d^3x \tag{5}$$

is called the *mass quadrupole moment tensor*, and is related to the *moment of inertia tensor*

$$J_{ij} = \int \rho(r^2\delta^{ij} - x^i x^j)\, d^3x \tag{6}$$

by

$$D_{ij} = J_{kk}\delta_{ij} - 3J_{ij}, \tag{7}$$

where $J_{kk} = J_{11} + J_{22} + J_{33}$. Notice that, by definition, the mass quadrupole moment tensor is traceless, $D_{kk} = D_{11} + D_{22} + D_{33} = 0$.

1.1.2 Calculate the mass quadrupole moment tensor of a homogeneous body having the shape of an ellipsoid.

Solution: Let the surface of the ellipsoid be given by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{1}$$

By introducing the new coordinates $x' = x/a$, $y' = y/b$, and $z' = z/c$, the volume integration over the ellipsoid reduces to that over the unit sphere.

Hence we have, for example,

$$\begin{aligned}
D_{11} &= \int\int\int \rho(3x^2 - r^2)\, dx\, dy\, dz \\
&= \int\int\int \rho(2x^2 - y^2 - z^2)\, dx\, dy\, dz \\
&= \int\int\int \rho abc\big(2a^2x'^2 - b^2y'^2 - c^2z'^2\big)\, dx'\, dy'\, dz' \\
&= \rho abc\big(2a^2 - b^2 - c^2\big)\int\int\int z'^2\, dx'\, dy'\, dz' \\
&= \rho abc\big(2a^2 - b^2 - c^2\big)\int_0^{2\pi}\int_0^{\pi}\int_0^1 r^4\, dr \cos^2\theta \sin\theta\, d\theta\, d\phi \\
&= \frac{m}{5}(2a^2 - b^2 - c^2), \qquad (2)
\end{aligned}$$

where $m = 4\pi abc\rho/3$ is the mass of the ellipsoid. Likewise, the other non-vanishing components of the mass quadrupole moment tensor are given by

$$D_{22} = \frac{m}{5}(-a^2 + 2b^2 - c^2) \qquad (3)$$

$$D_{33} = \frac{m}{5}(-a^2 - b^2 + 2c^2). \qquad (4)$$

1.1.3 Write the general term in the expansion of the Newtonian potential using spherical harmonics.

Solution: We expand the expression $1/|\mathbf{r} - \mathbf{r}'|$ into *spherical harmonics*:

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{\left(r^2 + r'^2 - 2rr'\cos\beta\right)^{1/2}} = \sum_{l=0}^{\infty} \frac{r'^l}{r^{l+1}} P_l(\cos\beta). \qquad (1)$$

Here β is the angle between vectors $\mathbf{r}$ and $\mathbf{r}'$ (for notation see Problem 1.1.1). Using now the addition theorem for the spherical harmonics, we obtain

$$P_l(\cos\beta) = \sum_{m=-l}^{l} \frac{(l - |m|)!}{(l + |m|)!} P_l^{|m|}(\cos\theta) P_l^{|m|}(\cos\theta') e^{-im(\phi - \phi')}, \qquad (2)$$

where the spherical angles θ, ϕ and θ', ϕ' denote the directions of the vectors $\mathbf{r}$ and $\mathbf{r}'$, respectively, with respect to the fixed coordinate system. The functions $P_l^m(\cos\theta)$ are the associated Legendre polynomials.

Introducing now the *spherical functions* defined by

$$Y_{lm}(\theta, \phi) = (-1)^{m_{i}l}\left[\frac{(2l+1)(l-m)!}{2(l+m)!}\right]^{1/2} P_l^m(\cos\theta)e^{im\phi}, \tag{3}$$

for $m \geqslant 0$, and

$$Y_{l,-|m|}(\theta, \phi) = (-1)^{l-m}\overline{Y}_{l,|m|}, \tag{4}$$

the expansion given above can then be written as

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\frac{r'^l}{r^{l+1}}Y_{lm}(\theta', \phi')\overline{Y}_{lm}(\theta, \phi). \tag{5}$$

If we now write the Newtonian potential in the form

$$\phi(x) = -G\int\frac{\rho(x')\,d^3x'}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty}\phi_l(x), \tag{6}$$

then the lth term will have the form

$$\phi_l = \frac{-G}{r^{l+1}}\sum_{m=-l}^{l}\left(\frac{4\pi}{2l+1}\right)^{1/2}Q_l^m\overline{Y}_{lm}(\theta, \phi), \tag{7}$$

where use has been made of the notation

$$Q_l^m = \left(\frac{4\pi}{2l+1}\right)^{1/2}\int\rho(x')r'^lY_{lm}(\theta', \phi')\,d^3x'. \tag{8}$$

The $2l+1$ quantities Q_l^m, with $m = -l, -l+1, \ldots, l$, describe the 2^l-pole moment of the mass system. The quantity $Q_0^0 = 2\pi^{1/2}m$, where m is the total mass of the system. The quantities Q_2^m, with $m = -2, -1, 0, 1, 2$, are related to the components of the mass quadrupole moment tensor D_{ij} by

$$\begin{aligned} Q_2^0 &= -\tfrac{1}{2}D_{33} \\ Q_2^{\pm 1} &= \pm\frac{1}{\sqrt{6}}(D_{13} \pm iD_{23}) \\ Q_2^{\pm 2} &= -\frac{1}{2\sqrt{6}}(D_{11} - D_{22} \pm 2iD_{12}). \end{aligned} \tag{9}$$

1.1.4 Solve the Laplace equation (1.1.5) in cylindrical coordinates, using the method of the separation of variables.

Solution: In cylindrical coordinates ρ, z, and ϕ, the Laplace equation takes the form

$$\frac{\partial^2 f}{\partial \rho^2}+\frac{1}{\rho}\frac{\partial f}{\partial \rho}+\frac{1}{\rho^2}\frac{\partial^2 f}{\partial \phi^2}+\frac{\partial^2 f}{\partial z^2}=0. \tag{1}$$

The separation of variables can then be achieved by the following substitution:

$$f(\rho, z, \phi)=R(\rho)Z(z)\Phi(\phi). \tag{2}$$

Using the solution (2) in the Laplace equation (1) then yields the following three differential equations:

$$\frac{d^2R}{d\rho^2}+\frac{1}{\rho}\frac{dR}{d\rho}+\left(k^2-\frac{\nu^2}{\rho^2}\right)R=0 \tag{3a}$$

$$\frac{d^2Z}{dz^2}-k^2Z=0 \tag{3b}$$

$$\frac{d^2\Phi}{d\phi^2}+\nu^2\Phi=0. \tag{3c}$$

Here k^2 and ν^2 are separation constants.

The solutions of the last two equations are elementary and are given by

$$Z(z)=e^{\pm kz} \tag{4a}$$

$$\Phi(\phi)=e^{\pm i\nu\phi}. \tag{4b}$$

In order that the potential f be single valued, ν must be an integer. The parameter k, on the other hand, is arbitrary and may be assumed to be real. By changing variables from ρ into $x=k\rho$, the radial equation (3a) becomes

$$\frac{d^2R}{dx^2}+\frac{1}{x}\frac{dR}{dx}+\left(1-\frac{\nu^2}{x^2}\right)R=0. \tag{5}$$

Equation (5) is the familiar *Bessel equation* whose solutions are *Bessel functions* of order ν.

We assume that the solution of the Bessel equation can be written in the form of a power series as

$$R(x)=x^{\alpha}\sum_{k=0}^{\infty}a_k x^k. \tag{6}$$

Then we find that $\alpha = \pm\nu$, and the coefficients a_k are given by

$$a_{2j-1} = 0 \tag{7a}$$

$$a_{2j} = -\frac{1}{4j(j+\alpha)}a_{2j-2}, \tag{7b}$$

for $j = 1, 2, 3, \ldots$. Hence the coefficients of the odd powers of x vanish. Iterating the recursion formula then yields

$$a_{2j} = \frac{(-1)^j\Gamma(\alpha+1)}{2^{2j}j!\Gamma(j+\alpha+1)}a_0. \tag{8}$$

If we choose the coefficient $a_0 = 1/2^\alpha\Gamma(\alpha+1)$, then the two solutions, corresponding to $\alpha = \pm\nu$, are given by

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{(-1)^j}{j!\Gamma(j+\nu+1)}\left(\frac{x}{2}\right)^{2j} \tag{9a}$$

$$J_{-\nu}(x) = \left(\frac{x}{2}\right)^{-\nu} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!\Gamma(j-\nu+1)}\left(\frac{x}{2}\right)^{2j}. \tag{9b}$$

These are *Bessel functions of the first kind* of order $\pm\nu$. The series converge for all finite values of x. If we assume that $\nu = m$ is an *integer*, the above two solutions are then linearly *dependent*, and we have

$$J_{-m}(x) = (-1)^m J_m(x). \tag{10}$$

If ν is taken to be *not* an integer, however, the two solutions $J_{\pm\nu}(\alpha)$ are then linearly *independent*.

We may replace the two solutions (9) by $J_\nu(x)$ and $N_\nu(x)$, where

$$N_\nu(x) = \frac{J_\nu(x)\cos\nu\pi - J_{-\nu}(x)}{\sin\nu\pi} \tag{11}$$

is a *Neumann function*, or a *Bessel function of the second kind*. The function $N_\nu(x)$ is linearly *independent* of $J_\nu(x)$, both when ν is not an integer and in the limit $\nu \to$ integer.

Finally, *Bessel functions of the third kind*, called *Hankel functions*, are defined as linear combinations of $J_\nu(x)$ and $N_\nu(x)$ by

$$H_\nu^{(1)}(x) = J_\nu(x) + iN_\nu(x) \tag{12a}$$

$$H_\nu^{(2)}(x) = J_\nu(x) - iN_\nu(x). \tag{12b}$$

Hankel functions also provide independent solutions to the Bessel equation, just as $J_\nu(x)$ and $N_\nu(x)$ do.

1.2 BASIC PROPERTIES OF THE GRAVITATIONAL FIELD

The first observation about gravitational phenomena is the fact that the attraction between bodies roughly follows an inverse-square law, and that it is very weak. This fact is well known from the study of planetary orbits. Indeed, the gravitational forces are very weak as compared to the other forces that exist between particles. Let us, for instance, calculate the ratio of the gravitational force to the electrical force between two electrons. The result is

$$\frac{F_{\text{gravitation}}}{F_{\text{electricity}}} = \frac{Gm_e^2}{e^2} = 0.24 \times 10^{-42}.$$

All other fields which we know are much stronger than the gravitational field.

Some of the basic properties of the gravitational field can be exhibited by merely observing the motion of particles in it. Such observations lead to the finding that all particles move in the same way in the gravitational field. The motion of particles occurs irrespective of the size of the mass of the particle.

When a particle moves in a gravitational field, then there are two kinds of forces which act on it: the inertial force and the gravitational force. Both of these forces, as has been shown in the last section, are proportional to the mass of the particle. By Newton's second law of motion the total sum of all forces acting on the particle equals zero. We therefore find that the mass terms of the particle cancel out from the equation of motion of the particle. This fact is a basic property of the gravitational field.

This basic property of gravitation also enables one to describe the motion of a particle in a gravitational field, but considered from a noninertial system of coordinates, and vice versa. For instance, the motion of a particle in a uniform gravitational field, as observed in an inertial coordinate system, may be looked upon as that of another particle which is not moving at all, but is being observed in terms of a noninertial coordinate system.

Figure 1.2.1 describes two physically completely different motions. Figure 1.2.1*a* shows a uniform gravitational field. A particle with mass m is subject to a force equal to $m\mathbf{g}$. Figure 1.2.1*b* shows no gravitational field, but the coordinate system is noninertial and moving along the z axis with an acceleration $\mathbf{a} = \mathbf{g}$. The inertial force, acting on the particle, is also equal to $m\mathbf{g}$.

It should be emphasized, however, that the equivalence of gravitational fields to noninertial fields of forces is local in character. It is not a complete equivalence in a global sense. One cannot always replace gravitational forces by inertial ones in an extended region of space.

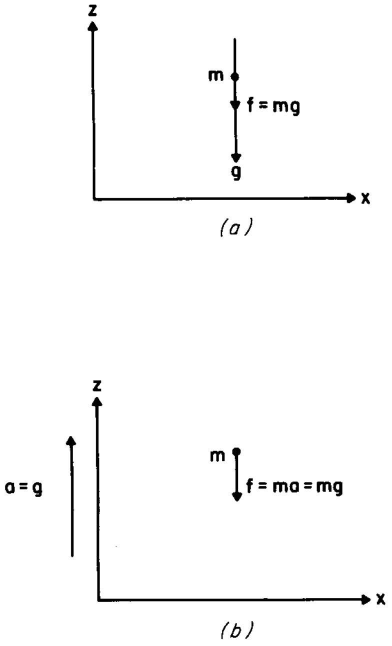

Figure 1.2.1 Equivalence of gravitational forces and inertial forces. (*a*) Particle located in uniform gravitational field with gravity pull equal to **g**. (*b*) Particle in noninertial coordinate system, moving along the z axis with an acceleration $\mathbf{a} = \mathbf{g}$, with no gravitational field. The forces acting on the two particles are equal.

The local equivalence between gravitational fields and fields of forces, which are due to noninertial systems of coordinates, may also be examined from a different point of view. By use of an inertial system we can always get rid of the noninertial effects that are equivalent to the gravitational field. We cannot eliminate a genuine gravitational field, however, in an extended region of space. We can eliminate the gravitational field only locally. This can be achieved by choosing a freely falling noninertial coordinate system in the gravitational field. Then locally, and only locally, the gravitational field appears to be eliminated.

If we extend our discussion to particles having relativistic velocities, there appears to be no change in the equivalence between gravitational fields and fields of inertial forces. The special relativistic line element, or proper time, when Cartesian coordinates are used, is given by

$$ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu = c^2\, dt^2 - dx^2 - dy^2 - dz^2, \tag{1.2.1}$$

where $\eta_{\mu\nu}$ is the flat-space Minkowskian metric,

$$\eta_{\mu\nu} = \begin{pmatrix} +1 & & & 0 \\ & -1 & & \\ & & -1 & \\ 0 & & & -1 \end{pmatrix}, \tag{1.2.2}$$

when an inertial system of coordinates is used. In the above equations, Greek indices μ, ν take the values $0, 1, 2, 3$, and repeated indices indicate the use of the summation convention. The constant c is the speed of light. The form of the special relativistic proper time does not change if one goes from one inertial system of coordinates into another by means of a Lorentz transformation.

Suppose now, however, that one goes from an inertial system into a noninertial system of coordinates. The square of the distance between two neighboring points ds^2 will no longer retain its Minkowskian form. When noninertial coordinate systems are used, ds^2 can no longer be written as the sum and difference of the squares of the four coordinate differentials. For instance, suppose that one goes into a uniformly rotating coordinate system. If one chooses the rotation to be around the z axis, then the transformation to the rotating coordinate system can be taken in the form

$$\begin{aligned} x &= x' \cos \omega t - y' \sin \omega t \\ y &= x' \sin \omega t + y' \cos \omega t \\ z &= z'. \end{aligned} \tag{1.2.3}$$

Here ω is the angular velocity of the rotation between the original inertial and the new noninertial coordinate systems.

When Eqs. (1.2.3) are used, the line element (1.2.1) will now have the form

$$\begin{aligned} ds^2 = {} & \left[c^2 - \omega^2\left(x'^2 + y'^2\right)\right] dt^2 + 2\omega\, dt\left(y'\, dx' - x'\, dy'\right) \\ & - \left(dx'^2 + dy'^2 + dz'^2\right). \end{aligned} \tag{1.2.4}$$

We therefore see that the line element is no longer the sum or the difference of the squares of the differentials of the coordinates.

In general when noninertial coordinate systems are used, the line element will include terms which are products of different coordinate differentials. As a result we have, in the general case, the following expression:

$$ds^2 = g_{\mu\nu}(x)\, dx^\mu\, dx^\nu, \tag{1.2.5}$$

where $g_{\mu\nu}(x)$ are ten functions of the space and time coordinates, with $g_{\mu\nu} = g_{\nu\mu}$. The coordinates are now being denoted by $x^0 = ct$, $x^1 = x$, $x^2 = y$, and $x^3 = z$. We may, furthermore, use non-Cartesian coordinates. In that case the coordinates x^1, x^2, and x^3 describe curvilinear coordinates. The ten functions $g_{\mu\nu}$ appear in the theory of Riemannian geometry (see next chapter), and play a very important role in that theory. They are called the components of the *metric tensor*.

As will be seen in the following chapters, the *general theory of relativity* assumes that the gravitational field, as well as fields of inertia, can all be described by means of the components of the metric tensor $g_{\mu\nu}$. From the theory of Riemannian geometry one knows that the determination of the metric tensor is equivalent to the determination of a four-dimensional curved spacetime. When the metric tensor $g_{\mu\nu}$ is obtained from the Minkowskian flat-space metric tensor $\eta_{\mu\nu}$, by going from an inertial coordinate system into a noninertial one, the transformation can be reversed. In that case we again obtain the special relativistic Minkowskian metric tensor. If the metric tensor $g_{\mu\nu}$ describes a genuine gravitational field, however, we cannot find a transformation that brings the line element (1.2.5) back into the special relativistic form (1.2.1). Only locally, in an infinitesimal region of the space, we can achieve that situation.

When such a situation occurs, the spacetime is called *curved* or *Riemannian*. A particular case of Riemannian spacetime is the special relativistic Minkowskian flat spacetime. While a Riemannian spacetime might have little or no symmetry whatsoever, the Minkowskian spacetime has actually the maximum number of degrees of symmetry, namely, ten degrees of freedom. These are the degrees of freedom of the familiar Poincaré group (nonhomogeneous Lorentz group). In this sense the *general* theory of relativity has only *special* symmetries, whereas the *special* theory of relativity has the most *general* symmetry of spacetime.

In the next section we discuss in more detail the experimental verification of the equality of inertial and gravitational masses.

1.3 NULL EXPERIMENTS

One of the most accurate experiments in physics is carried out to check whether or not there is a difference between the gravitational and the inertial masses of bodies. The experiment shows with great precision that all bodies fall in the gravitational field with the same acceleration. In this sense the experiment is actually a *null experiment*, since it fails to show any difference between the two kinds of masses. The experiment is usually referred to as the *Eötvös experiment*.

The experiment to verify the equality of the gravitational and inertial masses was apparently first performed by Galileo about the year 1610, and was repeated by Newton about the year 1680. Both Galileo and Newton have demonstrated experimentally that the acceleration of a body, in the gravity field of Earth, does not depend on the composition of the body. In both cases the experiment was performed using a pendulum. In 1827 the experiment was repeated by Bessel, also using a pendulum, but obtaining a more accurate result.

The experiment was subsequently repeated by Eötvös in the year 1890 using a torsion balance, verifying the same null result to an accuracy of about 10^{-8} for the ratio of the difference between the gravitational and the inertial masses, divided by the mass of the body. Some years later Eötvös, Pekar, and Fekete

Table 1.3.1 List of experiments showing evidence of the equality of the gravitational mass and the inertial mass.

Year	Experimenter(s)	Instrument	Accuracy[a]
1610	Galileo	Pendulum	2×10^{-3}
1680	Newton	Pendulum	2×10^{-3}
1827	Bessel	Pendulum	2×10^{-5}
1890	Eötvös	Torsion balance	5×10^{-8}
1922	Eötvös, Pekar, Fekete	Torsion balance	3×10^{-9}
1935	Renner	Torsion balance	2×10^{-10}
1964	Dicke, Roll, Krotkov	Torsion balance	3×10^{-11}
1971	Braginsky, Panov	Torsion balance	9×10^{-13}

[a] The accuracy is given by the expression $|m_G - m_I|/m_I$, where m_G and m_I are the gravitational and inertial masses, respectively.

repeated the experiment, improving the accuracy to about 10^{-9}. In 1935 Renner, also using a torsion balance, repeated the null experiment to an accuracy of 10^{-10}.

More recently the experiment was repeated by Dicke, and by Braginsky and Panov. These recent experiments were both done using torsion balances. The accuracies obtained were about 10^{-11} by Dicke and about 10^{-13} by Braginsky and Panov. From the above brief description of the results of the Eötvös experiment we obviously can conclude that the gravitational and inertial masses are indeed equal to each other to a great experimental precision. Table 1.3.1 summarizes some of the experimental evidence for the equality of the gravitational and inertial masses.

The experiment of Eötvös may be described as follows. Two pieces of matter are tied to the arms of a torsion balance. The two masses are made of different compositions. The first body is assumed to have an inertial mass m_I and a gravitational mass m_G, whereas the second body's inertial and gravitational masses are assumed to be given by m'_I and m'_G, respectively. The experiment then shows whether or not the ratios of the gravitational mass to the inertial mass are the same for both bodies. If these ratios are the same for both bodies, then no torque will apply on the balance. If the ratios m_G/m_I are different for the two bodies, however, then a new torque will be produced and will consequently apply on the balance, thus causing it to rotate.

Figure 1.3.1 shows the forces acting on the two bodies. Two forces act on each body: the gravitational force $m_G\mathbf{g}$, which is due to the Earth's pull to its center, and the centrifugal force $m_I\mathbf{a}$ produced by the rotation of the Earth around its axis. We assume that the beam of the balance points in the east–west direction. The centrifugal force is then directed along both the z axis and the x axis. Its components are therefore given by m_Ia_z and m_Ia_x, respectively, for the first body, for instance. The gravitational force $m_G\mathbf{g}$ is directed along the negative z axis.

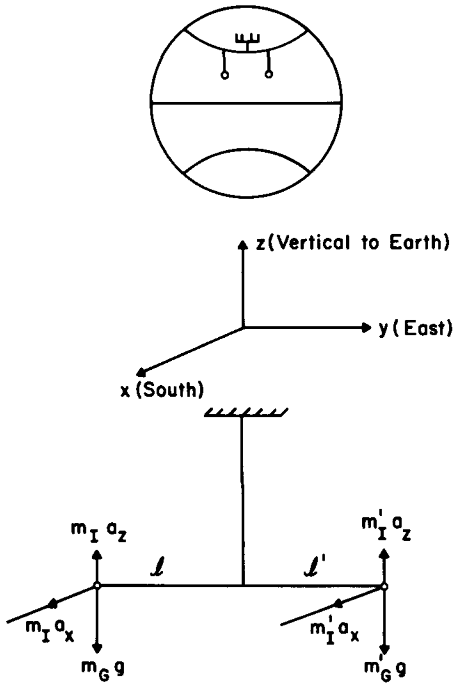

Figure 1.3.1 Schematic diagram of the Eötvös experiment. The experiment is performed with a torsion balance placed on the surface of the Earth at a latitude of approximately 45° north.

We can now calculate the components of the torque acting on the balance. The component of the torque along the z axis is given by

$$M_z = m_I a_x l - m'_I a_x l'. \tag{1.3.1}$$

The component of the torque along the x axis, on the other hand, is given by

$$M_x = (m_G g - m_I a_z)l - (m'_G g - m'_I a_z)l'. \tag{1.3.2}$$

This latter component of the torque, however, is equal to zero, $M_x = 0$, since the balance is in equilibrium condition. Eliminating l' from the above two equations, we obtain for the z component of the torque the following expression:

$$M_z = m_I a_x l \frac{m'_G/m'_I - m_G/m_I}{m'_G/m'_I - a_z/g}. \tag{1.3.3}$$

Equation (1.3.3) shows that a torque, directed along the z axis, will exist if the ratios of the gravitational mass to the inertial mass of the two bodies are different from each other. If M_z is different from zero, then the balance will rotate. As a result, the torque M_z will be balanced by an opposite torque produced by the suspension fiber of the balance. Hence no direct detection of the torque can be made. The presence of the torque M_z can be detected, however, by rotating the apparatus by 180° around the vertical axis. Such a rotation is equivalent to exchanging the masses of the two bodies m and m'. This, in turn, causes a change in the sign of the torque component M_z. Hence if the equilibrium position of the beam was along the east–west direction before the rotation, then it will be deviated from that direction after the rotation of the apparatus by 180° if there exists a nonvanishing torque component M_z.

The experiment by Dicke is also done using a torsion balance. The gravitational force used in this latter experiment, however, is that of the Sun rather than that of the Earth. The centrifugal force here is also different from that in the Eötvös experiment. It is the centrifugal force due to the Earth's motion around the Sun.

The Dicke experiment yields the following expression for the component of the torque in the vertical direction:

$$M_z = \left[(m_G g - m_I a)l - (m'_G g - m'_I a)l'\right]\sin\phi. \tag{1.3.4}$$

In the above equation we assume, for the sake of simplicity, that the experiment is being done at the north pole of the Earth, and therefore the centrifugal force is horizontal. Here g is the acceleration due to the gravity field of the Sun, a is the centrifugal acceleration at the Earth's location, and ϕ is the angle between the beam of the balance and the direction of the Sun. The beam is being held fixed at the laboratory and the torque is measured. The torque then oscillates with a period of 24 hours as the angle ϕ increases by 360°. In this way any nonperiodic effect can be isolated and eliminated. Actually the Dicke apparatus is more complicated than that of Eötvös, and the torsion balance uses three bodies rather than two.

In the next two sections we discuss the principles of the theory of general relativity, namely, the principle of equivalence and the principle of general covariance. We will see that the negative results of the Eötvös experiment provide convincing evidence supporting these principles.

1.4 PRINCIPLE OF EQUIVALENCE

In the last section we have seen that the Eötvös-type experiment shows, with a great accuracy, the equality of the gravitational mass and the inertial mass of bodies with different composition and structure of matter. We also found that the experiment gives the following relation between the two kinds of masses:

$$\frac{|m_G - m_I|}{m_I} \simeq 10^{-12}. \tag{1.4.1}$$

Thus the gravitational force acting on a body, which is proportional to the gravitational mass, has a certain relationship to the inertial force acting on the body, the latter force being proportional to the inertial mass. This relationship between the gravitational force and the inertial force, or the equivalence between them, is the basis of the *principle of equivalence.*

The principle of equivalence may be stated in many versions, one of which is as follows: *The gravitational forces and the inertial forces, acting on a body, are equivalent and are indistinguishable from each other.*

It is sometimes also convenient to make a distinction between two kinds of principles, a strong principle of equivalence and a *weak principle of equivalence.* The first is the principle upon which general relativity theory is based, and the latter is being supported directly by the Eötvös experiment.

The strong principle of equivalence might be stated in the following way: *In a freely falling, nonrotating, laboratory the local laws of physics take on some standard form, including a standard numerical content, independent of the position of the laboratory in space.* The weak principle of equivalence, on the other hand, says much less and might be stated as follows: *The local gravitational acceleration of a body is substantially independent of the composition and the structure of the matter of the body that is being accelerated.*

It is clear that the weak principle of equivalence is supported directly and strongly by the Eötvös-type experiment. It is not clear at all, however, that the Eötvös-type experiment supports also the strong principle of equivalence.

A consequence of the principle of equivalence, as stated above, is the fact that there cannot exist, within the framework of general relativity theory, inertial coordinate systems. Thus when gravitation exists in a certain region of space, then no inertial system can be introduced and used there. This is so since in such a system, by its very definition, all inertial accelerations vanish and, as a result, the gravitational accelerations can be isolated from the inertial accelerations. But this is in contradiction to the principle of equivalence. Only locally can we introduce inertial coordinate systems.

The impossibility of adapting inertial coordinate systems in general relativity theory, where the systems differ from each other by motion with constant velocities, makes the concept of acceleration no longer absolute as in Newtonian physics or in the special theory of relativity. In this sense, just as the special theory of relativity assumes that the velocity has a relative meaning only, the general theory of relativity makes the same assumption with respect to acceleration.

In the next section we discuss the second principle, the principle of general covariance, on which the theory of general relativity is based.

1.5 PRINCIPLE OF GENERAL COVARIANCE

As has been explained in the previous sections, the gravitational field may be eliminated locally by use of a freely falling coordinate system. By using a noninertial coordinate system, on the other hand, noninertial forces are obtained, and on a local basis they behave like gravitational forces and are indistinguishable from them. Hence we see that the special theory of relativity cannot be valid in an extended region of space where the gravitational field is presented. The special theory of relativity is valid locally only at each point of spacetime. In general, when gravitation is presented, however, a curved spacetime is needed to accommodate the gravitational field. Moreover, all laws of physics should be generally covariant under the most general coordinate transformations.

In the original formulation of the theory of general relativity by Einstein, two principles were advocated as a basis for the theory. These are the principle of equivalence discussed in the last section and the *principle of general covariance*.

The principle of general covariance can be stated in one of the following forms:

1 All coordinate systems are equally good for stating the laws of physics. Hence all coordinate systems should be treated on the same footing, too.
2 The equations that describe the laws of physics should have tensorial forms and be expressed in a four-dimensional Riemannian spacetime.
3 The equations describing the laws of physics should have the same form in all coordinate systems.

As can be seen, the three versions for describing the principle of general covariance are not quite equivalent. From the above description, however, we learn that according to the principle of general covariance the coordinates in general relativity theory are nothing more than a bookkeeping system to label the spacetime events. Physical consequences and results should not depend on the particular coordinate system used to obtain the results. Only quantities that are invariant under the choice of the coordinate system will usually have physical significance.

The principle of general covariance is usually a valuable guide to deriving the equations for the laws of physics. In particular, the principle is most useful when we try to generalize the laws of physics from their special relativistic form into their general relativistic form, thus incorporating the gravitational field into these equations. This use of the principle of general covariance is repeatedly made in general relativity theory. It is similar to the use of the *principle of minimal coupling* in field theory, but it is much more powerful.

For example, the Maxwell equations, when coupled with gravitation, have to be generalized into curved spacetime (see Section 3.4). Application of the principle of general covariance then shows how the Maxwell equations may be generalized. The generalized Maxwell equations should be such that they go back to the ordinary flat-space equations when gravitation is turned off. In this procedure we mostly, but not necessarily always, replace the usual partial derivatives by covariant derivatives.

In conclusion, while the principle of equivalence necessarily leads to the introduction of a curved spacetime, the principle of general covariance guides us in formulating the equations for the laws of physics. In fact, these two principles are most of what is needed to generate Einstein's theory of general relativity, which is said to be the greatest single achievement of theoretical physics. They lead directly to the idea that the gravitational phenomena can be beautifully described and accommodated by means of the Riemannian geometry of spacetime.

With the above remarks we end our preliminary introduction to gravitational phenomena. In the next chapter we develop the mathematical tools needed to describe general relativity theory and gravitation.

SUGGESTED REFERENCES

J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics*, McGraw-Hill, New York, 1965. (Section 1.5)

V. B. Braginsky and A. B. Menukin, *Measurement of Weak Forces in Physics Experiment*, University of Chicago Press, Chicago, 1977. (Section 1.3)

V. B. Braginsky and V. I. Panov, *Zh. Eksp. Teor. Fiz.* **61**, 873 (1971). [English translation: *Sov. Phys.—JETP* 34, 463 (1972).] (Section 1.3)

M. Carmeli, *Group Theory and General Relativity*, McGraw-Hill, New York, 1977. (Section 1.5)

R. H. Dicke, *Sci. Am.* **205**(6) 84 (1961). (Section 1.3)

R. H. Dicke, *Theoretical Significance of Experimental Relativity*, Gordon and Breach, New York, 1964. (Section 1.4)

R. V. Eötvös, *Math. Natur. Berg. Ungarn* **8**, 65 (1890); *Beil. Ann. Phys.* **15**, 688 (1891); *Ann. Phys.* **59**, 354 (1896). (Section 1.3)

R. V. Eötvös, D. Pekar, and E. Fekete, *Ann. Phys.* **68**, 11 (1922). (Section 1.3)

J. Renner, *Hung. Acad. Sci.* **53**, 542 (1935). (Section 1.3)

P. G. Roll, R. Krotkov, and R. H. Dicke, *Ann. Phys.* (*N.Y.*) **26**, 442 (1964). (Section 1.3)

A. Trautman, *Lectures on General Relativity*, Prentice-Hall, Englewood Cliffs, NJ, 1965. (Section 1.5)