CHAPTER

14

The Role of Relativity in Stellar Structure and Gravitational Collapse

In Chap. 6 we discussed the metric of the exterior of a spherically symmetric distribution of mass, and in Chap. 7 we extended our considerations to include rotation. In this chapter we shall investigate the interior of massive bodies. For simplicity we limit ourselves to spherically symmetric systems with the special energy-momentum tensor (10.41), i.e., perfect fluids.

Our considerations have application in the study of stars which have reached the final stages of evolution. During most of the life of a star the light nuclei in the interior combine, i.e., undergo fusion, to release large quantities of energy. Much of this is in the form of radiation. This radiation produces a pressure that helps to counter the inward force of gravity, thereby stabilizing the star. For stars in which the fusion process has nearly ceased and little radiation pressure remains we may reasonably expect the stellar material to be approximately described by a perfect-fluid energy-momentum tensor in which phenomena such as viscosity and heat conduction are ignored. Such material, no longer capable of significant energy release via fusion, is generally referred to as *cold catalyzed matter:* it is cold in the sense that it behaves thermodynamically like a zero-temperature fluid and catalyzed in the sense that the fusion energy has nearly all been extracted.

After initial considerations on the basic equations of relativistic stellar structure for cold catalyzed matter we discuss the simple model of Schwarzschild, in which the proper density ρ is a constant. This will be followed by a discussion of the stability properties of very dense stars of cold catalyzed matter, which leads naturally to questions on the evolution of such stars. The simplest example of gravitational collapse, the spherical dust ball, will then be treated.

Throughout this chapter our purpose is to illustrate the role played by general relativity in astrophysics and not to do realistic calculations, since these are usually very involved and not as enlightening as more simplified examples.

14.1 Relativistic Stellar Structure

In this section we shall set up the problem of stellar structure in terms of the perfect-fluid energy-momentum tensor representing cold catalyzed stellar material. This involves the construction of a suitable form for the metric and the statement of its relation to the density and pressure inside the star. This must be combined with a study of the physical interpretation of a number of mathematical statements that emerge. For example, one of the most important of these equations we obtain will be a generalization of the Newtonian equation of hydrodynamic equilibrium, known as the Tolman-Oppenheimer-Volkov (TOV) equation.

At all times we shall assume that we are dealing with a static and spherically symmetric configuration of mass, in which the density ρ and the pressure p are functions of only a radial coordinate r:

$$\rho = \rho(r) \qquad p = p(r) \tag{14.1}$$

As we shall discuss further, a local relation is usually assumed to exist between these quantities; this is called the *equation of state*, and may be written as

$$p = p(\rho) \tag{14.2}$$

In Sec. 6.1 we have already discussed the general form of a metric which is static and spherically symmetric. We showed that we can bring it into the form

$$ds^2 = e^{\nu}c^2\,dt^2 - [e^{\lambda}\,dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\varphi^2)] \tag{14.3}$$

The coordinates used may reasonably be identified with the familiar polar coordinates used with flat space; $\nu(r)$ and $\lambda(r)$ are functions of the radial variable and must be determined from the field equations. Since we are now dealing with different field equations than in Chap. 6, the determination of these functions will differ from that for the empty-space Schwarzschild problem.

The energy-momentum tensor which enters the field equations (10.101) was determined in (10.41). It is convenient to use lower indices, and

so we now write it as

$$T_{\alpha\beta} = \rho u_\alpha u_\beta + \frac{p}{c^2}(u_\alpha u_\beta - g_{\alpha\beta}) \tag{14.4}$$

Since the matter is at rest at each point, the components of the velocity four-vector u^α are $(u^0,0,0,0)$. On the trajectory of each particle of matter in the fluid the relation between proper-time and coordinate-time is given by

$$ds^2 = g_{00}(dx^0)^2 = g_{00}c^2\,dt^2 \qquad 1 = g_{00}(u^0)^2 \tag{14.5}$$

We have, furthermore,

$$u_0 = g_{0\alpha}u^\alpha = g_{00}u^0 = \sqrt{g_{00}} \qquad u_i = 0 \tag{14.6}$$

This allows us to write $T_{\alpha\beta}$ in the form

$$T_{\alpha\beta} = \rho\begin{pmatrix} g_{00} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} - \frac{p}{c^2}\begin{pmatrix} 0 & g_{01} & g_{02} & g_{03} \\ g_{10} & g_{11} & g_{12} & g_{13} \\ g_{20} & g_{21} & g_{22} & g_{23} \\ g_{30} & g_{31} & g_{32} & g_{33} \end{pmatrix} \tag{14.7}$$

which simplifies, by virtue of (14.2), to

$$T_{\alpha\beta} = \begin{pmatrix} \rho e^{\nu} & 0 & 0 & 0 \\ 0 & \frac{p}{c^2}e^{\lambda} & 0 & 0 \\ 0 & 0 & \frac{p}{c^2}r^2 & 0 \\ 0 & 0 & 0 & \frac{p}{c^2}r^2\sin^2\theta \end{pmatrix} \tag{14.8}$$

for the ideal fluid at rest. When we insert this into the field equations (10.101), we shall naturally obtain a set of relations between the geometric functions $\nu(r)$ and $\lambda(r)$ and the fluid parameters $\rho(r)$ and $p(r)$. The form of the field equations (10.101*b*) is the most convenient for our purpose because the scalar $T_\alpha{}^\alpha = T$ is easily obtained from (14.4), namely

$$T = \rho - \frac{3p}{c^2} \tag{14.9}$$

which follows from $u^\alpha u_\alpha = 1$ and $g_\alpha{}^\alpha = 4$. We therefore have all the nonzero terms of the right side of the field equations (10.101) and may write

$$
\begin{aligned}
T_{00} - \tfrac{1}{2}g_{00}T &= \frac{e^\nu}{2}\left(\rho + \frac{3p}{c^2}\right)\\
T_{11} - \tfrac{1}{2}g_{11}T &= \frac{e^\lambda}{2}\left(\rho - \frac{p}{c^2}\right)\\
T_{22} - \tfrac{1}{2}g_{22}T &= \frac{1}{2}\left(\rho - \frac{p}{c^2}\right) r^2\\
T_{33} - \tfrac{1}{2}g_{33}T &= \frac{1}{2}\left(\rho - \frac{p}{c^2}\right) r^2 \sin^2\theta
\end{aligned}
\tag{14.10}
$$

To get the left side of the field equations in terms of $\nu(r)$ and $\lambda(r)$, we first need the contracted Riemann tensor, or Ricci tensor, $R_{\mu\lambda}$. This is, fortunately, a very easy task in the present case, for we have already obtained all the components of $R_{\mu\lambda}$ associated with the metric tensor (14.2) in our discussion of the Schwarzschild solution in Sec. 6.1. Indeed, referring back to (6.31), (6.35), (6.44), and (6.49), we see that the nonzero components of $R_{\mu\nu}$ are

$$
\begin{aligned}
R_{00} &= e^{\nu-\lambda}\left[-\frac{\nu''}{2} + \frac{\nu'\lambda'}{4} - \frac{\nu'^2}{4} - \frac{\nu'}{r}\right]\\
R_{11} &= \frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} - \frac{\lambda'}{r}\\
R_{22} &= e^{-\lambda}\left[1 + \frac{\nu' r}{2} - \frac{\lambda' r}{2}\right] - 1\\
R_{33} &= R_{22}\sin^2\theta
\end{aligned}
\tag{14.11}
$$

where the prime denotes differentiation with respect to r. The field equations are thus

$$e^{-\lambda}\left[-\frac{\nu''}{2} + \frac{\nu'\lambda'}{4} - \frac{\nu'^2}{4} - \frac{\nu'}{r}\right] = C\left[\frac{\rho}{2} + \frac{3p}{2c^2}\right] \tag{14.12a}$$

$$e^{-\lambda}\left[\frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} - \frac{\lambda'}{r}\right] = C\left[\frac{\rho}{2} - \frac{p}{2c^2}\right] \tag{14.12b}$$

$$e^{-\lambda}\left[\frac{1}{r^2} + \frac{\nu' - \lambda'}{2r}\right] - \frac{1}{r^2} = C\left[\frac{\rho}{2} - \frac{p}{2c^2}\right] \tag{14.12c}$$

where $C = -8\pi\kappa/c^2$. Note that we now have only three equations since the equation involving R_{33} is clearly proportional to that involving R_{22}. These can be put into more convenient form. We first add (14.12*a*) and (14.12*b*) to get

$$-C\left[\rho + \frac{p}{c^2}\right] = e^{-\lambda}\left(\frac{\lambda' + \nu'}{r}\right) \tag{14.13}$$

Note that since C is negative and the density and pressure are greater than or equal to zero, this implies that $\lambda' + \nu'$ is positive or zero and equal to zero only for free space, that is, $\rho = p = 0$. We can now solve (14.12*c*) and (14.13) for ρ and p; for a third equation we eliminate ρ and p from (14.12*b*) and (14.12*c*). This gives the simple system

$$C\rho = e^{-\lambda}\left(\frac{1}{r^2} - \frac{\lambda'}{r}\right) - \frac{1}{r^2} \tag{14.14a}$$

$$C\frac{p}{c^2} = \frac{1}{r^2} - e^{-\lambda}\left(\frac{1}{r^2} + \frac{\nu'}{r}\right) \tag{14.14b}$$

$$\frac{e^\lambda}{r^2} = \frac{1}{r^2} - \frac{\nu'^2}{4} + \frac{\nu'\lambda'}{4} + \frac{\nu' + \lambda'}{2r} - \frac{\nu''}{2} \tag{14.14c}$$

Up to this point the density and pressure of the fluid have been treated as arbitrary independent scalar functions. The system (14.14) represents three ordinary differential equations for the four functions which describe the geometry and the physics of the system, namely, $\lambda(r)$, $\nu(r)$ and $p(r)$, $\rho(r)$. We still have the mathematical freedom to make arbitrary assumptions on the physical nature of the fluid which constitutes the system. This is usually done by prescribing a pressure-density relation, the equation of state $p = p(\rho)$ of the fluid.

Let us digress for a moment, however, to indicate an interesting alternative way of obtaining stellar models. We may prescribe e^ν arbitrarily inside the star, asking only that it be equal to $1 - 2m/r$ at the boundary in order that it match the Schwarzschild exterior solution there. Then (14.14*c*) is a simple first-order differential equation that can be solved for λ by quadratures. Then we can compute $p(r)$ and $\rho(r)$ from (14.14*a*) and (14.14*b*) and determine an equation of state $p(\rho)$. It only remains to be checked whether the solution is physically reasonable or not, in particular if p and ρ are positive inside the star (Adler, 1974).

Let us now return to the problem of solving the system (14.14) in the case where an equation of state is prescribed. In recent years a great deal of effort has gone into obtaining equations of state for cold catalyzed matter up to and even beyond nuclear densities, i.e., roughly 10^{14} g/cm^3.

As we shall discuss further in Sec. 14.3, these equations of state are considered to be relatively trustworthy despite the extreme densities involved.

In analogy with the Schwarzschild solution of Chap. 6, let us first define a function $m(r)$ by

$$e^{-\lambda} = 1 - \frac{2m(r)}{r} \tag{14.15}$$

This function $m(r)$ can be shown to play the role of the geometric mass $\kappa M/c^2$ inside a sphere of radius r. To see this observe that

$$-\frac{1}{r^2}(2m'(r)) = -\frac{1}{r^2}[r(1-e^{-\lambda})]' = e^{-\lambda}\left(\frac{1}{r^2} - \frac{\lambda'}{r}\right) - \frac{1}{r^2} \tag{14.16}$$

From (14.14*a*) it then follows that

$$m'(r) = \frac{\kappa\rho}{c^2}(4\pi r^2) \tag{14.17}$$

or

$$m(r) = \int_0^m dm = \frac{\kappa}{c^2}\int_0^r 4\pi r^2\rho\, dr \tag{14.18}$$

where we have set $m = 0$ at $r = 0$ in order to avoid a zero in the metric term e^{λ} at $r = 0$. Equation (14.18) justifies the interpretation of $m(r)$ as the geometric mass inside radius r. Outside the boundary of the star we may use (14.15) with m equal to the *total mass* of the star; this of course is the exterior Schwarzschild solution of g_{11} that we discussed in Chap. 6. Thus we automatically have a continuous metric function g_{11} across the stellar surface.

With the use of the function $m(r)$ we can obtain a very useful equation for p', the derivative of p with respect to r, in terms of ρ, p, m, and r. Observe first that (14.14*b*) can be solved for ν' in a convenient form

$$\nu' = 2\,\frac{m + 4\pi\kappa p r^3/c^4}{r(r-2m)} \tag{14.19}$$

Next, we relate p' to ν'. Differentiate (14.14*b*) and utilize (14.14*c*) to eliminate ν'' from the result

$$\begin{aligned} -\frac{8\pi\kappa}{c^4}p' &= -\frac{2}{r^3} + e^{-\lambda}\left[\frac{\lambda'}{r^2} + \frac{\lambda'\nu'}{r} - \frac{\nu''}{r} + \frac{\nu'}{r^2} + \frac{2}{r^3}\right] \\ &= e^{-\lambda}(\nu' + \lambda')\frac{\nu'}{2r} \end{aligned} \tag{14.20}$$

Comparing this with (14.13), we obtain a simple relation between p' and ν':

$$\frac{p'}{c^2} = -\frac{\nu'}{2}\left(\rho + \frac{p}{c^2}\right) \tag{14.21}$$

If the expression (14.19) is substituted for ν', we obtain finally

$$p' = -\frac{(\rho + p/c^2)(m + 4\pi\kappa p r^3/c^4)c^2}{r(r-2m)} \tag{14.22}$$

This is the famous TOV equation.

In Sec. 14.3 we shall discuss the use of the TOV equation in the construction of stellar models. Let us now show that the TOV equation is the relativistic generalization of the Newtonian equation of hydrostatic equilibrium. Consider, in the context of classical theory, a small rectangular box of fluid in the model star; the bottom is at a radial distance r and has area dA, while the top is at a radial distance $r + dr$ and also has an area dA. The net upward force on the element due to the pressure differential is easily seen to be $-p'\,dr\,dA$. The condition of hydrostatic equilibrium is that the Newtonian gravitational force exerted on this element by the rest of the star must balance the force due to pressure. The downward gravitational force is simply

$$F = \frac{\kappa M(r)\rho\, dA\, dr}{r^2} \tag{14.23}$$

where $M(r)$ is the total Newtonian mass inside r. Thus equilibrium ensues if

$$p' = \frac{-\kappa M(r)\rho}{r^2} \tag{14.24}$$

If we identify $m(r) = \kappa M(r)/c^2$, as above, we see that this is just the limit of the TOV equation for $r \gg 2m$, $\rho \gg p/c^2$, and $m \gg 4\pi\kappa p r^3/c^4$; in most normal stars the pressure and density are low enough for these limits to represent very reasonable approximations, and Eq. (14.24) may be used in place of (14.22). However, many neutron stars are sufficiently dense to require use of the TOV equation.

Let us now collect in summary the main results of this section, (14.15), (14.17), (14.21), (14.22), which together with the equation of state we shall refer to as the equations of relativistic stellar structure for static cold catalyzed matter:

(14.25a) $$p = p(\rho) \qquad \text{(the equation of state)}$$

(14.25b) $$m' = \frac{4\pi\kappa\rho r^2}{c^2}$$

(14.25c) $$\frac{p'}{c^2} = -\frac{m + 4\pi\kappa p r^3/c^4}{r(r - 2m)}\left(\rho + \frac{p}{c^2}\right)$$

(14.25d) $$e^{-\lambda} = 1 - \frac{2m}{r}$$

(14.25e) $$\nu' = -\frac{2p'}{\rho c^2 + p}$$

The first three equations form a simple first-order system that can in principle be solved to yield functions $m(r)$, $\rho(r)$, and $p(r)$ if initial conditions are given, for example, $m(0) = 0$ and central density $\rho(0) = \rho_c$. The radius of the model star is naturally taken to be that r_0 for which the pressure vanishes, $p(r_0) = 0$, and the total mass is $m(r_0)$. We shall always assume that such a radius exists. From $m(r)$ one finds the metric function e^λ from (14.25d). To obtain the remaining metric function ν it is necessary to solve (14.25e); the solution will be arbitrary up to a constant, which can be determined by matching the interior solution to the exterior Schwarzschild solution, $e^{\nu(r_0)} = 1 - 2m(r_0)/r_0$.

The solution of the system (14.25) will be carried through for a special case in Sec. 14.2 and discussed further in 14.3.

14.2 A Simple Stellar Model—The Interior Schwarzschild Solution

We wish to solve the system (14.25) in all details for the very simple case of constant $\rho(r)$. This model was discussed by Schwarzschild and recommends itself primarily by its great mathematical simplicity. A constant density ρ does not imply that the physical fluid density must be constant, since the physical density depends on the metric, which is not constant. This will be discussed further in the latter part of this section. For constant ρ we can integrate (14.25b) immediately to obtain

(14.26) $$m(r) = \frac{4\pi\kappa\rho r^3}{3c^2}$$

and therefore, by (14.25d),

(14.27) $$e^{-\lambda} = 1 - \frac{8\pi\kappa\rho r^2}{3c^2}$$

For notational convenience let us define a quantity $\hat{R}$ with the dimensions of a length by the equation

(14.28) $$\hat{R}^2 = \frac{3c^2}{8\pi\kappa\rho}$$

We can then write g_{11} in the very simple form

(14.29) $$g_{11} = -e^\lambda = -\left(1 - \frac{r^2}{\hat{R}^2}\right)^{-1}$$

The present case is somewhat artificial for a star and corresponds to the classical notion of an incompressible fluid: we have no equation of state giving p as a function of ρ. However, with constant ρ we can integrate at once the relation (14.25e) or (14.21) between pressure, density, and the metric function $\nu(r)$. We obtain

(14.30) $$\frac{8\pi\kappa}{c^2}\left(\rho + \frac{p}{c^2}\right) = De^{-\nu/2}$$

where D is an arbitrary constant of integration. This can be substituted into (14.13) to yield a differential relation between $\lambda(r)$ and $\nu(r)$:

(14.31) $$\frac{e^{-\lambda}}{r}(\nu' + \lambda') = De^{-\nu/2}$$

Since $e^{-\lambda}$ is known from the preceding paragraph, we now have a differential equation for $\nu(r)$. In order to solve it we rearrange (14.31) to

(14.32) $$rDe^{-\nu/2} = e^{-\lambda}\nu' - (e^{-\lambda})'$$

Substituting for $e^{-\lambda}$ from (14.29) we then obtain

(14.33) $$rDe^{-\nu/2} = \left(1 - \frac{r^2}{\hat{R}^2}\right)\nu' + \frac{2r}{\hat{R}^2}$$

To solve this differential equation let $e^{\nu/2} = \gamma(r)$; since $\gamma'(r) = (\nu'/2)e^{\nu/2}$, we can bring (14.33) into the final form

$$\left(1 - \frac{r^2}{\hat{R}^2}\right)\gamma'(r) + \frac{r}{\hat{R}^2}\gamma = \tfrac{1}{2}rD \tag{14.34}$$

We guess at once an obvious particular solution of this inhomogeneous differential equation; namely,

$$\gamma_p = \tfrac{1}{2}D\hat{R}^2 \tag{14.35}$$

On the other hand, the corresponding homogeneous differential equation

$$\left(1 - \frac{r^2}{\hat{R}^2}\right)u'(r) + \frac{r}{\hat{R}^2}u(r) = 0 \tag{14.36}$$

has the general solution

$$u(r) = B\left(1 - \frac{r^2}{\hat{R}^2}\right)^{1/2} \tag{14.37}$$

Thus the function $\gamma(r) = e^{\nu/2}$ must have the form

$$e^{\nu/2} = \tfrac{1}{2}D\hat{R}^2 - B\left(1 - \frac{r^2}{\hat{R}^2}\right)^{1/2} \tag{14.38}$$

with a properly chosen constant of integration B. We have thus determined the last unknown component of the metric tensor,

$$g_{00} = e^{\nu} = \left[\tfrac{1}{2}D\hat{R}^2 - B\left(1 - \frac{r^2}{\hat{R}^2}\right)^{1/2}\right]^2 \tag{14.39}$$

If we denote by

$$A = \tfrac{1}{2}D\hat{R}^2 \qquad D = \frac{2\rho}{3}\frac{8\pi\kappa}{c^2}A \tag{14.40}$$

a new constant of integration, we can express the Schwarzschild line element in the interior of the sphere of fluid as follows. Using (14.29) and (14.39), we have

$$ds^2 = \left[A - B\left(1 - \frac{r^2}{\hat{R}^2}\right)^{1/2}\right]^2 c^2\,dt^2 - \left(1 - \frac{r^2}{\hat{R}^2}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \tag{14.41}$$

It remains only to require zero pressure at the stellar surface and to fuse this line element with the exterior Schwarzschild solution of Chap. 6. This is very easily done with the coordinates and conventions we have set up.

Let us note before continuing that the interior line element develops a singularity in g_{11} when $r = \hat{R}$, which is reminiscent of the behavior of the exterior Schwarzschild solution discussed in Chap. 6. For the present we shall suppose that the radius of the star is $r_0 < \hat{R}$. Later we discuss the situation where r_0 approaches the critical value $\hat{R}$.

To determine A we demand that the pressure in the fluid be zero at the surface, and thereby join continuously with the zero pressure of space outside the fluid. If we substitute the expression (14.38) and (14.40) into the relation (14.30), we obtain an expression for $p(r)$

$$\left(\rho + \frac{p}{c^2}\right) = \frac{2\rho A/3}{A - B(1 - r^2/\hat{R}^2)^{1/2}} \tag{14.42}$$

The demand that $p = 0$ at $r = r_0$ leads to

$$1 = \frac{2A/3}{A - B(1 - r^2/\hat{R}^2)^{1/2}} \tag{14.43}$$

or

$$A = 3B\left(1 - \frac{r_0^2}{\hat{R}^2}\right)^{1/2} \tag{14.44}$$

so that only the arbitrary constant B remains in the metric.

To evaluate B we demand that the metric functions g_{00} and g_{11} join continuously with the exterior Schwarzschild metric functions. Thus, assuming $r_0 > 2m$,

$$\begin{aligned} 1 - \frac{2m}{r_0} &= 1 - \frac{r_0^2}{\hat{R}^2} \qquad m = \frac{\kappa M}{c^2} \\ 1 - \frac{2m}{r_0} &= \left[A - B\left(1 - \frac{r_0^2}{\hat{R}^2}\right)^{1/2}\right]^2 = 4B^2\left(1 - \frac{r_0^2}{\hat{R}^2}\right) \end{aligned} \tag{14.45}$$

where we have used (14.44). This yields $B = \frac{1}{2}$ (the sign is arbitrary) and a relation between ρ and M that follows from the definition (14.28) of $\hat{R}^2$

$$M = \frac{4\pi}{3}r_0^3\rho \tag{14.46}$$

We now have determined the line element inside and outside the Schwarzschild model of a star:

$$(14.47)\quad ds^2 = \left[\frac{3}{2}\sqrt{1-\frac{r_0^2}{\hat{R}^2}} - \frac{1}{2}\sqrt{1-\frac{r^2}{\hat{R}^2}}\right]^2 c^2\,dt^2 - \frac{dr^2}{1-r^2/\hat{R}^2}$$

$$- r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \qquad \text{for } r \le r_0,\ \hat{R}^2 = \frac{3c^2}{8\pi\kappa\rho}$$

$$ds^2 = \left(1-\frac{2m}{r}\right)c^2\,dt^2 - \frac{dr^2}{1-2m/r}$$

$$- r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \qquad \text{for } r \ge r_0,\ m = \frac{\kappa M}{c^2}$$

We have so far supposed that the radius r_0 of the model star is greater than the Schwarzschild radius $2m$ so that no metric singularity occurs for $r > r_0$. This restricts the mass M accordingly

$$(14.48)\quad r_0 > 2m = \frac{2\kappa M}{c^2} \qquad M < \frac{c^2 r_0}{2\kappa}$$

We have also assumed that r_0 is less than $\hat{R}$, the parameter introduced in (14.28), so that no metric singularity occurs for $r < r_0$; it is interesting to note that this leads to the same result as above, (14.48). To see this we substitute into

$$(14.49)\quad r_0^2 < \hat{R}^2 = \frac{3c^2}{8\pi\kappa\rho}$$

the expression for ρ obtained from (14.46)

$$(14.50)\quad \rho = \frac{3M}{4\pi r_0^3}$$

to obtain again (14.48).

A slightly more stringent condition on M can be obtained from the pressure equation (14.42). If the pressure is never to become infinite inside the fluid, the denominator of (14.45) must never vanish. This will be so if $A > B$, or

$$(14.51)\quad \frac{3}{2}\left[1-\frac{r_0^2}{\hat{R}^2}\right]^{1/2} > \frac{1}{2}$$

The square of this relation yields

$$(14.52)\quad r_0^2 < \tfrac{8}{9}\hat{R}^2 = \frac{c^2}{3\pi\kappa\rho}$$

so by substituting for ρ from (14.50), we obtain, finally,

$$(14.53)\quad M < \frac{4}{9}\frac{c^2 r_0}{\kappa}$$

which is only slightly smaller than the previously imposed limit (14.48). Note also that (14.53) guarantees that the coefficient of $(dx^0)^2$ in the line element is positive even at the center of the sphere.

The relations (14.48) and (14.53) limit the mass of a sphere of fixed radius; alternatively, by the use of Eq. (14.50), they can be converted into a limit on the mass of a sphere of arbitrary radius but fixed ρ. Solving Eq. (14.50) for r_0, we have

$$(14.54)\quad r_0 = \left(\frac{3M}{4\pi\rho}\right)^{1/3}$$

which, inserted in (14.48), gives

$$(14.55)\quad M < \frac{c^2}{2\kappa}\left(\frac{3M}{4\pi\rho}\right)^{1/3} \qquad M^2 < \frac{3c^6}{32\pi\rho\kappa^3}$$

A similar procedure applied to (14.52) gives the more stringent limit

$$(14.56)\quad M^2 < \frac{16c^6}{243\pi\rho\kappa^3}$$

Next, as promised earlier, we study the problem of the physical interpretation of the assumption of constant ρ.

As discussed in Chaps. 3, 4, and 12, the physical volume element of a space with metric determinant g is $\sqrt{|g|}$ times the product of the coordinate intervals. Thus the physical three-dimensional volume element inside the Schwarzschild model star is

$$(14.57)\quad dV = \left(1-\frac{r^2}{\hat{R}^2}\right)^{-1/2} r^2 \sin\theta\,d\theta\,d\varphi\,dr$$

which differs from the corresponding classical volume element $r^2 \sin\theta$

$d\theta\, d\varphi\, dr$ by the factor $(1 - r^2/\hat{R}^2)^{-1/2}$, which is *greater than* 1. Equation (14.57) can easily be integrated to obtain the total volume V,

$$V = \int \frac{r^2 \sin\theta\, d\theta\, d\varphi\, dr}{[1 - r^2/\hat{R}^2]^{1/2}} = 4\pi \int_0^{r_0} \frac{r^2\, dr}{[1 - r^2/\hat{R}^2]^{1/2}} \tag{14.58}$$

The most convenient way to perform the final step of the integration is to define $\sin\alpha = r/\hat{R}$ and $\sin\alpha_0 = r_0/\hat{R}$, which yields

$$\begin{aligned} V &= 4\pi\hat{R}^3 \int_0^{\alpha_0} \sin^2\alpha\, d\alpha = 2\pi\hat{R}^3(\alpha_0 - \sin\alpha_0 \cos\alpha_0) \\ &= 2\pi\hat{R}^3 \left[\arcsin\frac{r_0}{\hat{R}} - \frac{r_0}{\hat{R}}\left(1 - \frac{r_0^2}{\hat{R}^2}\right)^{1/2} \right] \end{aligned} \tag{14.59}$$

For most normal stars $r_0/\hat{R}$ will be a very small number, and so we shall expand the parentheses in (14.59) in a power series in $r_0/\hat{R}$; using the well-known series for arcsine and square roots, we obtain

$$\left[\arcsin\frac{r_0}{\hat{R}} - \frac{r_0}{\hat{R}}\left(1 - \frac{r_0^2}{\hat{R}^2}\right)^{1/2} \right] = \frac{2}{3}\left(\frac{r_0}{\hat{R}}\right)^3 + \frac{1}{5}\left(\frac{r_0}{\hat{R}}\right)^5 + O\left(\left(\frac{r_0}{\hat{R}}\right)^7\right) \tag{14.60}$$

which, inserted in (14.59), yields

$$V = \frac{4\pi}{3} r_0^3 \left[1 + \frac{3}{10}\left(\frac{r_0}{\hat{R}}\right)^2 + O\left(\left(\frac{r_0}{\hat{R}}\right)^4\right) \right] \tag{14.61}$$

Thus, finally, the average density $\langle\rho\rangle$ of the sphere is

$$\langle\rho\rangle = \frac{M}{V} = \frac{3M}{4\pi r_0^3}\left[1 - \frac{3}{10}\left(\frac{r_0}{\hat{R}}\right)^2 + O\left(\left(\frac{r_0}{\hat{R}}\right)^4\right) \right] \tag{14.62}$$

If we substitute from (14.50), this can be expressed in terms of ρ as

$$\langle\rho\rangle = \rho\left[1 - \frac{3}{10}\left(\frac{r_0}{\hat{R}}\right)^2 + O\left(\left(\frac{r_0}{\hat{R}}\right)^4\right) \right] \tag{14.63}$$

That is, the average density and ρ differ by terms of order $(r_0/\hat{R})^2$. It is thus clear that ρ does *not* represent a constant "physical" density, for if it did, $\langle\rho\rangle$ would certainly be equal to ρ. Instead, we see that because of the curvature of space, via the factor $(1 - r^2/\hat{R}^2)^{-1/2}$, $\langle\rho\rangle$ differs from ρ, the difference becoming negligible for small values of $r_0/\hat{R}$.

The formula (14.63) may be interpreted in an interesting manner. We should expect a volume V with local density ρ to have the mass $V\rho$. We have a smaller mass M and may discuss the mass defect

$$V\rho - M = V\rho\left[\frac{3}{10}\left(\frac{r_0}{\hat{R}}\right)^2 + O\left(\left(\frac{r_0}{\hat{R}}\right)^4\right) \right] \tag{14.64}$$

We may attribute this mass defect to the loss of energy in packing the matter under its own gravitational energy. In classical mechanics a sphere of radius r and homogeneous density ρ has on its surface the potential

$$V(r) = -\kappa \frac{4\pi}{3}\frac{r^3\rho}{r} = -\frac{4\pi\kappa}{3} r^2\rho \tag{14.65}$$

If we increase the radius by dr, we bring the amount of matter $4\pi r^2\rho\, dr$ from the zero level of energy to the level $V(r)$ and lose the energy

$$dE(r) = -\frac{16\pi^2\kappa\rho^2}{3} r^4\, dr \tag{14.66}$$

Hence the energy loss in spherical packing due to gravitation is

$$E = \int_0^{r_0} dE = -\tfrac{16}{15}\pi^2\kappa\rho^2 r_0^5 \tag{14.67}$$

On the other hand, the mass defect (14.64) can be calculated to first approximation to be

$$\Delta M = \left(\frac{4\pi}{3} r_0^3\right)\left(\frac{8\pi\kappa\rho}{3c^2}\right)\frac{3r_0^2\rho}{10} = \tfrac{16}{15}\pi^2\kappa\rho^2 \frac{r_0^5}{c^2} \tag{14.68}$$

Thus

$$\Delta M = -\frac{E}{c^2} \tag{14.69}$$

and the mass defect appears accounted for by Einstein's fundamental mass-energy relation to the order of approximation used. It is therefore fundamental to note that the mass M which appears in the Schwarzschild metric represents all mass-energy contained in the source, even the negative gravitational binding energy.

14.3 Stellar Models and Stability

We have carried through the interior Schwarzschild solution in detail because it is mathematically simple yet illustrates the role of relativity in constructing a stellar model. Moreover, it serves to clarify features of the general problem, such as the physical interpretation of the density scalar ρ. It is desirable, however, to use a realistic equation of state in the study of white dwarf and neutron stars. Such stars are believed to be created in the aftermath of the cataclysmic explosion of a red giant star into a supernova. Great extremes of density may occur in the red giant core which remain after such an explosion; in white dwarfs the density is around 10^6 g/cm^3 and in neutron stars it approaches 10^{16} g/cm^3, about 100 times nuclear density. Despite the difficulty of dealing with such extreme densities, much work has been done on equations of state for cold catalyzed matter and many stellar models constructed. We wish to discuss qualitative features of these models.

With an equation of state and a given central density ρ_c, Eqs. (14.25*b*) and (14.25*c*) allow one to construct an interesting function, the mass of the star in terms of its central density. For this purpose it is necessary to integrate these equations for the functions p, ρ, and m, which will depend on the central density ρ_c. For an acceptable equation of state there must be some r_0 for which $p(r_0) = 0$, as we have already noted. This stellar radius and the total stellar mass $m(r_0)$ will naturally depend on ρ_c. We may thus think of the total stellar mass as a function $m(\rho_c)$ of ρ_c. In practice this function is usually obtained numerically with high-speed computers. It has very interesting qualitative features common to most realistic equations of state. We first consider models of white dwarf stars. On the basis of atomic physics an equation of state appropriate to the densely packed atoms of a white dwarf is obtainable. With such an equation of state it is found that the total mass increases monotonically for ρ_c in the interval of 10^5 g/cm^3 to about 10^9 g/cm^3, and reaches a maximum value of about 1.2 solar masses. This maximum mass is referred to as the *Chandrasekhar limit* (see Fig. 14.1). A stellar model with a mass greater than the Chandrasekhar limit is unstable and must turn into a time-dependent system. Thus theory predicts that no stable white dwarf stars with a mass greater than about 1.2 solar masses can exist, which is confirmed by all observations to date. This result is not critically dependent on the details of the equation of state used since the dominant pressure is produced by a so-called degenerate electron gas, the physics of which is rather well understood. The result is also independent of general relativity in that the classical limit (14.24) of the TOV equation may be used as an excellent approximation.

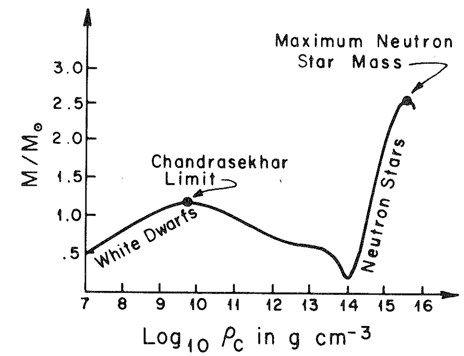
For masses greater than the Chandrasekhar limit it is necessary to

reconsider the problem as time-dependent. The model star is found to contract, and extreme pressures and densities occur in the interior. This causes electrons to be absorbed by protons to produce neutrons and neutrinos, a process known as *inverse beta decay*. By this process a star of great density and small size, about 10 km, composed largely of neutrons can form. Another region of stability occurs on the curve in Fig. 14.1 for such neutron stars. Pulsars, sources which have been observed by astronomers to emit intense bursts of electromagnetic energy in precisely timed pulses, have been identified as neutron stars in rapid rotation.

Neutron stars can be studied in the same manner as white dwarfs. The main differences are that the equation of state for a very dense gas of neutrons is not as well understood as for the electron gas, and relativistic effects are important. Nevertheless one qualitative feature appears to be common to most models based on reasonable equations of state. The function $m(\rho_c)$ increases monotonically to a value of order one solar mass at a density of about 10^{15} g/cm^3 (see Fig. 14.1). Beyond this critical mass there is no known stable state for a superdense star; thus it is predicted that a stable neutron star will not exist with a mass very much greater than a solar mass. A heavier star would become unstable and shrink with time in the process known as gravitational collapse. It is extraordinary that the pressure obtained from most equations of state cannot balance the gravitational force and halt the collapse.

We have so far limited our discussion to white dwarf and neutron stars, which are limited in mass as noted. Another very interesting case of instability involves not great densities but very large masses. To see how this comes about consider the limit on the mass of a Schwarzschild

Fig. 14.1
Central density versus mass of white dwarf and neutron star models, showing the maximum values of the masses; see also Cohen and Cameron (1971).



star imposed by the demand that the pressure remain finite at the center (14.56)

$$M^2 < \frac{16c^6}{243\pi\rho\kappa^3}$$

For ρ of the order of 1 g/cm³, the density of water or of the sun, the critical value of M is about 10^8 solar masses. Thus we are led to expect instability to occur at very ordinary densities for a star of sufficiently great mass.

In the next section, we shall discuss the simplest example of the evolution of an unstable mass, the gravitational collapse of a spherical ball of dust with no internal pressure.

14.4 Gravitational Collapse of a Dust Ball

In the supernova explosion of a red giant star the small dense core of the red giant is left behind, often to become a white dwarf or neutron star. However, if this core is much more massive than a solar mass, there is no stable state, as we have discussed, and the core must collapse. Studies indicate that many red giant cores may be expected to exceed the stability limit. Since a realistic description of the collapse of such a stellar model would take us beyond the aims of this book, we shall consider a very simple mathematically tractable model; since the pressure generated during collapse is not adequate to halt the collapse, we shall take the drastic step of ignoring the pressure entirely in order to gain insight into the behavior of the geometry associated with a collapsing body. Thus, we shall study the collapse of a spherically symmetric dust ball, falling freely inward upon itself. This model is particularly simple since we can form it by piecing together previous results from Chaps. 6 and 13.

Let us consider the metric appropriate to a spherically symmetric ball of dust, of radius r_d and with uniform but time-dependent density ρ. The exterior metric is the Schwarzschild metric of Chap. 6, as may be inferred from the Birkhoff theorem mentioned in Chap. 6. The interior metric is relatively easy to obtain by using the results of Chaps. 12 and 13. Indeed, we shall adopt the Robertson-Walker metric and assume that the dust is at rest in a co-moving coordinate system, just as in the cosmological problem. The only difference is that in the present problem the co-moving radial coordinate u extends only up to some finite value u_d corresponding to the dust-ball radius instead of ranging over the whole of the universe, i.e. to infinity. The radius u_d is co-moving with the dust and is therefore taken to be *independent of time*. The physical

motion of the dust-ball surface is described by the function R, which is time-dependent. It is thus clear that except for the decreased range of the radial coordinate u the mathematical results of Chaps. 12 and 13 may be used for the dust-ball problem. For maximum simplicity we shall consider the case where the parameters k and Λ are both zero. Thus, in the absence of pressure, the metric and the equations governing R become, from (6.53) and (12.56):

Exterior:

$$ds^2 = \left(1 - \frac{2m}{r}\right) c^2\,dt^2 - \frac{dr^2}{1 - 2m/r} - r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \tag{14.70a}$$

Interior:

$$ds^2 = c^2\,d\tau^2 - R(\tau)[du^2 + u^2(d\theta^2 + \sin^2\theta\,d\varphi^2)] \tag{14.70b}$$

and from (13.18)

$$\frac{8\pi\kappa}{c^2}\rho = \frac{3R'^2}{R^2c^2} \qquad \frac{4\pi\kappa}{c^2}\rho = -\frac{3R''}{c^2R} \tag{14.71}$$

We have denoted the Robertson-Walker time coordinate by τ to distinguish it from the Schwarzschild time coordinate t.

Equations (14.70) and (14.71) set up the problem by describing a "truncated Friedman universe" joined at some radius to the exterior Schwarzschild solution. It now remains to relate the radial markers u and r, to solve for $R(\tau)$, and to verify that the metric forms in (14.70) join smoothly at the dust-ball radius, which we may denote by either u_d or r_d.

The connection between the radial coordinates r and u is easily obtained. Recall from Chap. 6 that the Schwarzschild radial coordinate is distinguished in the sense that the invariant area of a sphere of radius r is $4\pi r^2$, as in flat space. This is evident from the form of the angular part of the line element (14.70a); it is also clear from (14.70b) and the discussion following (12.71) that Ru is distinguished in the same way. We therefore identify the time-dependent physical radius of the dust cloud in Schwarzschild and co-moving Robertson-Walker coordinates as

$$r_d(t) = R(\tau)u_d \tag{14.72}$$

It is not necessary to solve here for $R(\tau)$ since we have already done this in Sec. 13.3. There we obtained the first integral

$$R'^2 = \frac{D_0}{R} c^2 \tag{14.73}$$

where D_0 is a constant of integration. Thus, the first integral for $u_d R(\tau)$ is

$$(u_d R')^2 = \frac{D_0 u_d^3}{u_d R} c^2 \tag{14.74}$$

which has the solution

$$(u_d R)^{3/2} = (u_d R(0))^{3/2} - \tfrac{3}{2} \sqrt{D_0 u_d^3}\, c\tau \tag{14.75}$$

We have chosen a minus sign for the second term; a plus sign would correspond to an exploding dust ball (Exercise 14.5).

This solution has several important features. First, the Robertson-Walker time coordinate is the same as the proper time, so that $c\tau$ and s are interchangeable in Eqs. (14.74) and (14.75). Second, it is evident that R' is never zero unless R is infinite, as may be seen from Eq. (14.73). Thus our solution cannot represent a dust ball of finite extent collapsing from rest; we must assume that R is infinite for a time in the infinite past. This is a consequence of using the value $k = 0$ and is the price we must pay for choosing the simplest case. It could be avoided by choosing $k = 1$, but we would then find a zero radius in the finite past (see Exercise 14.7).

Preparatory to verifying that the interior and exterior solutions in (14.71) match we give a physical interpretation to the constant D_0 in (14.75). Equations (13.18) specify D_0 as given in (13.23) as

$$D_0 = \left(\frac{4\pi}{3} R^3 \rho\right) \frac{2\kappa}{c^2} \tag{14.76}$$

We can show that $u_d^3 D_0$ is twice the total geometric mass of the dust ball as follows. The invariant three-volume element corresponding to the Robertson-Walker metric is

$$dV = \sqrt{-g}\, du\, d\theta\, d\varphi = R^3 u^2 \sin\theta\, du\, d\theta\, d\varphi \tag{14.77}$$

and so the mass of the cloud (see Sec. 13.1) is

$$M = \rho V = 4\pi \int_0^{u_d} \rho R^3 u^2\, du = \frac{4\pi}{3} u_d^3 R^3 \rho \tag{14.78}$$

since the three-space is Euclidean. Twice its geometric mass is

$$2m = \frac{2\kappa}{c^2} M = \frac{2\kappa}{c^2} \left(\frac{4\pi}{3} R^3 \rho\right) u_d^3 = u_d^3 D_0 \tag{14.79}$$

With $u_d^3 D_0$ identified as $2m$ we can rewrite (14.74) and (14.75) as

$$(u_d R')^2 = \frac{2mc^2}{u_d R} \tag{14.80a}$$

$$(u_d R)^{3/2} = (u_d R(0))^{3/2} - \tfrac{3}{2} \sqrt{2m}\, \tau \tag{14.80b}$$

It now remains only to verify that the interior and exterior solutions do indeed match at $r_d = u_d R(\tau)$. The picture which we obtain for the collapse of the ball is that its radius measured by $u_d R(\tau)$ decreases with time according to (14.80). Outside this radius we should have the Schwarzschild solution, (14.70*a*). Certainly if we have a consistent solution, the exterior metric is the Schwarzschild metric as guaranteed by the Birkhoff theorem, discussed in Chap. 6. We can be assured that the interior and exterior solutions must join if the motion of a test particle just inside the ball's radius agrees with that of a test particle just outside. The motion of a radially falling particle just inside the ball is governed by (14.80). That of a particle just outside the dust ball is governed by the geodesic equation for radial motion in a Schwarzschild field. Referring to Eq. (6.82) yields

$$\dot{r}_d^2 = \left(\frac{dr_d}{ds}\right)^2 = \frac{2m}{r_d} \tag{14.81}$$

where we have set the constant $h = 0$ for radial motion and have chosen $l = 1/c$ so that $\dot{r}_d = 0$ at $r_d = \infty$, as with the interior motion. This equation is the same as (14.80*a*). Our task is now completed, because r_d has been identified with $u_d R(\tau)$ and $c\tau$ has been identified with the proper time s, so that (14.80) and (14.81) tell us that the radial motion of a particle just inside the dust-ball surface is the same as one just outside the surface; i.e., it is the same in the two different geometries.

The properties of the dust-ball collapse as viewed from the outside are now evident. One would see the surface falling freely in precisely the same manner as the test particle discussed in Sec. 6.7. That is, it would shrink asymptotically to $r_d = 2m$, the black-hole radius. On the other hand, from the viewpoint of an observer moving with the dust-ball surface, the collapse would proceed to zero radius in a finite proper time and no singularities at all would occur in the metric. If we had constructed our dust-ball model with the choice $k = \pm 1$ instead of $k = 0$, we would have reached the same qualitative conclusion (see Exercise 14.6).

The dust ball is an unrealistic model for the collapse of a very dense star of normal size. However, as we noted in Sec. 14.3, one may expect instability also for bodies of great mass but very modest densities. To investigate this in the present context we consider (14.78) with the dust-ball radius $u_d R$ set equal to the asymptotic collapse radius $2m = 2\kappa M/c^2$. We then obtain a relation between M^2 and ρ in the asymptotic state of collapse

$$M^2 = \frac{3}{32\pi}\left(\frac{c^2}{\kappa}\right)^3 \frac{1}{\rho} \tag{14.82}$$

in direct analogy with (14.55). We may choose ρ to be sufficiently small so that if a realistic equation of state were being used, instead of $p = 0$, we could still expect p to be small and have negligible effect on the collapse. For example, if ρ were about 10^{-4} g/cm^3, collapse would still occur if M were about 10^{10} solar masses, the size of a small galaxy; at this very low density it would appear reasonable to neglect pressure. It is possible to speculate that very large conglomerates of gas and dust may also condense to form large low-density systems collapsing asymptotically to black holes.

Exercises

14.1 Solve the equations of stellar structure (14.25) for an ideal isothermal Boltzmann gas, which has an equation of state $p = \alpha c^2 \rho$, where α is a constant. To do this assume a solution of the form $\rho = Ar^n$ and determine the constants A and n.

14.2 (*continued*) The above solution is badly behaved for $r = 0$. Give a physical interpretation of this result. If a sphere of incompressible fluid, ρ = const, is placed at the center of the gas, the solution can be made well-behaved at the origin. Show this explicitly.

14.3 (*continued*) The above solution is also badly behaved in the sense that the pressure is nonzero for any finite radius. Discuss how this defect could be remedied by making the pressure drop to zero at some value of r chosen as the stellar surface.

14.4 (*continued*) One way to remedy the defect discussed above is to place a shell of constant density fluid around the gas. Demonstrate this. What mass must such a shell have?

14.5 Study an exploding dust ball by choosing a plus sign for the second term in (14.75).

14.6 A Schwarzschild exterior solution can be joined to a dust-ball interior solution for the cases $k = 1$ and $k = -1$ in an analogous manner to the $k = 0$ case discussed in Sec. 14.4. Do this calculation.

14.7 In Exercise 14.6 with $k = 1$ an infinite dust-ball radius at $t = -\infty$ is avoided, but this undesirable feature is replaced by a *zero* radius at some finite past time. Show and discuss this.

14.8 Obtain a solution for a dust-filled universe containing a spherically symmetric cavity in which a spherically symmetric body is placed. Do this by joining a standard Schwarzschild solution to a Robertson-Walker solution extending from $u = u_d$ to $u = \infty$. Consider all cases of k.

14.9 Following the above exercises, join a dust ball of radius u_d to a Schwarzschild solution, and the Schwarzschild solution to a spherically symmetric dust cloud of inner radius $u = u_i > u_d$ and extending to $u = \infty$.

14.10 (*continued*) Discuss the evolution of this system and its physical interpretation in terms of an idealized collapsing star in an otherwise dust-filled universe.

14.11 Study the behavior of collapsing and exploding dust balls using classical Newtonian mechanics, and compare with the results of relativity theory.

14.12 What are the Petrov types of the metric of the collpasing dust-ball problem in various regions of space-time?

Problems

14.1 The effects of slow rotation may be added to the stellar structure problem in a simple manner. Begin this study by obtaining a reasonable metric for a slowly rotating fluid body, working to first order in the rotation rate (see Adams et al., 1973).

14.2 (*continued*) Obtain the Einstein equations for a slowly rotating fluid body, allowing the rotation rate to be a function of position in the body. Again work to first order in the rotation rate.

14.3 (*continued*) Solve the Einstein equations for the gaseous model, considered in Exercises 14.1 to 14.4, when it is rotating slowly. The problem is simplified if the shell thickness is allowed to go to a zero limit.

14.4 (*continued*) Solve the Einstein equations for a slowly rotating incompressible fluid, as in the Schwarzschild interior solution.

14.5 Study the red shift of light emitted radially from the surface of a collapsing dust ball and also the exploding dust ball of Exercise 14.5.

14.6 Obtain the wave equation for sound propagation in a perfect fluid, and show that the speed of sound is $(dp/d\rho)^{1/2}$.

14.7 For the speed of sound not to exceed c we must demand that $(dp/d\rho) < c^2$. Thus for the ideal isothermal Boltzmann gas α must be less than 1. Is this restriction physically reasonable?

Bibliography

Adams, R. C., J. M. Cohen, R. J. Adler, and C. Sheffield (1973): Analytic Neutron Star Models, *Phys. Rev.*, **D8**:1651.

Adler, R. J. (1974): A Fluid Sphere in General Relativity, *J. Math. Phys.*, **15**:727.

Carter, B. (1971): Axisymmetric Black Hole Has Only Two Degrees of Freedom, *Phys. Rev. Letters*, **26**:331.

Cohen, J. M., and A. G. W. Cameron (1971): Neutron Star Models, Including the Effects of Hyperon Formation, *Astrophys. Space Sci.*, **10**:227.

Israel, W. (1967): Event Horizons in Static Vacuum Space-Times, *Phys. Rev.*, **164**: 1776.

Oppenheimer, J. R., and H. Snyder (1939): On Continued Gravitational Contraction, *Phys. Rev.*, **56**:455.

Ruffini, R., and J. A. Wheeler (1971): Relativistic Cosmology and Space Platforms, *Proc. Conf. Space Physics, ESRO Paris Meeting.*

Schwarzschild, K. (1916): Über das Gravitationsfeld einer Kugel aus inkompressibler Flüssigkeit nach der Einsteinschen Theorie, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 424–434.

CHAPTER

15

Electromagnetism and General Relativity

General relativity is primarily a theory of gravitation. It enables us to understand the rather mysterious physical "force" of gravitation in terms of the purely geometric structure of the space-time manifold. Gravitation, however, is not the only force that occurs in classical physics. Electromagnetic forces are as universal and important as gravitational forces, and their agent, the electromagnetic field, is not explained by classical general relativity as a geometric phenomenon. Thus there have been many attempts to imbed the theory of the electromagnetic field into the framework of an extended theory of general relativity.

The ideas of Weyl (1918, 1922) and Eddington (1923) are particularly interesting. These authors attempt to introduce electromagnetic potentials as geometric quantities which determine the law of transplantation of a scale of length between different points and the comparison of length units in different directions at the same point. One associates in this way the electromagnetic potential with some sort of length distortion in space-time. We shall discuss this attempt to geometrize the electromagnetic field in a brief sketch in Sec. 15.2.

Einstein (1955) devoted much research in the later years of his life to a unified theory of the gravitational and the electromagnetic field which should describe both in terms of the metric tensor. For this purpose he had to assume the metric tensor to be nonsymmetric; he thereby obtained just a sufficient number of new field variables to describe the electromagnetic field. This unified field theory has been worked out in great mathematical detail by Hlavaty (1957), to whose book the reader is referred for further information on this theory.

However, despite the efforts of the physicists named above and many other ingenious attempts, it can safely be stated that no unified theory of electromagnetism and gravitation has been developed which is as convincing and satisfactory as Einstein's original theory of the gravitational field alone. This is most unfortunate since many physicists feel that the best classical description of the elementary particles is that of singularities in a combined electromagnetic-gravitational field.

In this chapter we shall pursue a more modest problem. In Sec. 15.1 we shall solve the combined Einstein-Maxwell equations in the standard classical form for the simplest case of physical interest, namely, a charged mass point. In the final sections of this chapter we shall then study briefly the ideas of Wheeler and Misner et al. for an "already unified" field theory based on the combined Maxwell and Einstein field equations (Einstein, 1955; Misner and Wheeler, 1957; Wheeler, 1957, 1961, 1962).

15.1 The Field of a Charged Mass Point

We shall first study the field of a charged mass point, that is, a point singularity of the Einstein field equations with an energy-momentum tensor (10.70) due to an electromagnetic field. We shall assume both the metric and the electromagnetic field to be spherically symmetric and time-independent. Such a situation represents the simplest example of a combined gravitational-electromagnetic field with sufficient physical significance. The problem of determining the coupled fields is somewhat similar to the interior Schwarzschild problem of Sec. 14.2, except that the energy-momentum tensor is now due to the electric field of a point charge instead of describing a fluid sphere. Observe that in the present case, because of the antisymmetry of the electromagnetic field tensor, the electromagnetic energy-momentum tensor (10.69) has zero trace. Hence, we can simplify the general field equations (10.101) to the form

$$R_{\mu\nu} = CT_{\mu\nu} \qquad C = -\frac{8\pi\kappa}{c^2} \tag{15.1}$$

We must solve this system together with the classical free-space Maxwell equations, which we shall write in the form given in (4.70*a*) and (4.70*b*). Thus the system to be solved is

$$R_{\mu\tau} = CT_{\mu\tau} = \frac{C}{c^2}[F_{\mu\alpha}F^{\alpha}{}_{\tau} + \tfrac{1}{4}g_{\mu\tau}F_{\alpha\beta}F^{\alpha\beta}] \tag{15.2a}$$

$$(\mathfrak{F}^{\mu\tau})_{|\tau} = (\sqrt{-g}\,F^{\mu\tau})_{|\tau} = 0 \tag{15.2b}$$

$$\{F_{\mu\tau|\lambda}\} = 0 \tag{15.2c}$$

There must, of course, be a singularity allowed in both the $F_{\mu\tau}$ and $g_{\mu\tau}$ fields at the position of the particle, or one will obtain the identically vanishing solution. We assume the particle at the origin of our polar coordinate system; clearly it is also the center of symmetry.

Since we are dealing with a Schwarzschild-type particle, the symmetry considerations which led to the general form of the Schwarzschild metric tensor in Chaps. 6 and 14 apply here as well. Thus the $g_{\mu\nu}$ and $g^{\mu\nu}$ will again have the form

$$g_{\mu\tau} = \begin{pmatrix} e^{\nu} & 0 & 0 & 0 \\ 0 & -e^{\lambda} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2\sin^2\theta \end{pmatrix} \tag{15.3}$$

$$g^{\mu\tau} = \begin{pmatrix} e^{-\nu} & 0 & 0 & 0 \\ 0 & -e^{-\lambda} & 0 & 0 \\ 0 & 0 & -r^{-2} & 0 \\ 0 & 0 & 0 & -\dfrac{1}{r^2\sin^2\theta} \end{pmatrix}$$

where ν and λ are functions of r, which must be obtained from the field equations. Furthermore, the tensor $R_{\mu\tau}$, which is constructed entirely from $g_{\mu\tau}$, may be carried over intact from Chaps. 6 and 14. From (14.11) we therefore obtain

$$\begin{aligned} R_{00} &= e^{\nu-\lambda}\left[-\frac{\nu''}{2} + \frac{\lambda'\nu'}{4} - \frac{\nu'^2}{4} - \frac{\nu'}{r}\right] \\ R_{11} &= \frac{\nu''}{2} - \frac{\lambda'\nu'}{4} + \frac{\nu'^2}{4} - \frac{\lambda'}{r} \\ R_{22} &= e^{-\lambda}\left[1 + \frac{r\nu'}{2} - \frac{r\lambda'}{2}\right] - 1 \\ R_{33} &= R_{22}\sin^2\theta \\ R_{\mu\tau} &= 0 \qquad \text{for } \mu \neq \tau \end{aligned} \tag{15.4}$$

where a prime denotes differentiation with respect to r. The major labor of the problem, working out $R_{\mu\tau}$ from $g_{\mu\tau}$, is thus already done.

The Maxwell tensor $F_{\mu\tau}$ of the problem should correspond to a static and spherically symmetric electric field $E(r)$ in the r, that is, x^1 direction. From the form of $F_{\mu\tau}$ in special relativity,

$$F_{\mu\tau} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & H_z & -H_y \\ E_y & -H_z & 0 & H_x \\ E_z & H_y & -H_x & 0 \end{pmatrix} \qquad \text{(special relativity)} \tag{15.5}$$

we are therefore led to seek a solution with $F_{\mu\tau}$ in the form

$$F_{\mu\tau} = E(r)\begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{15.6}$$

Notice that, since $F_{\mu\tau}$ is a function of r only, this form of $F_{\mu\tau}$ automatically satisfies the second Maxwell equation (15.2c) independent of the function $E(r)$. $E(r)$ must be determined along with ν and λ from (15.2a) and (15.2b). Raising indices in (15.6) with the $g^{\mu\tau}$ in (15.3), we easily obtain

$$F^{\mu\tau} = e^{-(\nu+\lambda)}E\begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{15.7}$$

and

$$\mathfrak{F}^{\mu\tau} = \sqrt{-g}\,F^{\mu\tau} = e^{-(\nu+\lambda)/2}r^2E\sin\theta\begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{15.8}$$

Maxwell's equation (15.2b) can now be solved for E. We substitute (15.8) into (15.2b) to obtain the only nontrivial condition,

$$\mathfrak{F}^{01}{}_{|1} = [e^{-(\nu+\lambda)/2}r^2E\sin\theta]' = 0 \tag{15.9}$$

Thus

$$e^{-(\nu+\lambda)/2}r^2E = \epsilon \qquad \epsilon = \text{const} \tag{15.10}$$

or

$$E = e^{(\nu+\lambda)/2}\frac{\epsilon}{r^2} \tag{15.11}$$

This is an explicit solution for the field E in terms of the yet unknown functions ν and λ. The boundary condition that the geometry be Euclidean at infinity implies that ν and λ approach zero as $r \to \infty$, so the solution (15.11) has the usual classical form, at least for large r. Thus the constant ϵ *can be identified as the charge of the particle.*

In order to solve the remaining equation (15.2a), we need to compute the $T_{\mu\tau}$ of the electric field. We use the solution (15.11) and the metric

tensor (15.3) to write the Maxwell tensor (15.6) in its three possible forms:

$$F_{\mu\tau} = e^{(\nu+\lambda)/2}\frac{\epsilon}{r^2}\begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{15.12a}$$

$$F^{\mu\tau} = e^{-(\nu+\lambda)/2}\frac{\epsilon}{r^2}\begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{15.12b}$$

$$F_\mu{}^\tau = \frac{\epsilon}{r^2}\begin{pmatrix} 0 & e^{(\nu-\lambda)/2} & 0 & 0 \\ e^{(\lambda-\nu)/2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{15.12c}$$

An elementary calculation then gives the result that

$$T_{\mu\tau} = \frac{1}{c^2}\left[F_{\mu\alpha}F^\alpha{}_\tau + \tfrac{1}{4}g_{\mu\tau}F_{\alpha\beta}F^{\alpha\beta}\right] = \frac{\epsilon^2}{2c^2r^4}\begin{pmatrix} e^\nu & 0 & 0 & 0 \\ 0 & -e^\lambda & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix} \tag{15.13}$$

We can now proceed to solve (15.2a), the Einstein equations. Using R_{00} and R_{11} from (15.4) and T_{00} and T_{11} from (15.13), we write the first two Einstein equations as

$$e^{\nu-\lambda}\left[-\frac{\nu''}{2} + \frac{\lambda'\nu'}{4} - \frac{\nu'^2}{4} - \frac{\nu'}{r}\right] = \frac{C\epsilon^2}{2c^2r^4}e^\nu \tag{15.14a}$$

$$\frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} - \frac{\lambda'}{r} = -\frac{C\epsilon^2}{2c^2r^4}e^\lambda \tag{15.14b}$$

Now multiply the first (15.14a) by $e^{-(\nu-\lambda)}$ and add to the second (15.14b) to get

$$\lambda' + \nu' = 0 \qquad \lambda + \nu = \text{const} \tag{15.15}$$

Since both λ and ν approach zero by the boundary condition at $r = \infty$, the constant in (15.15) must be zero, and we see that

$$\lambda = -\nu \tag{15.16}$$

This is the same relation that occurs in the ordinary Schwarzschild solution.

The remaining equation $R_{22} = CT_{22}$ is obtained from (15.4) and (15.13) as

$$e^{-\lambda}[1 + \tfrac{1}{2}r(\nu' - \lambda')] - 1 = \frac{C\epsilon^2}{2c^2r^2} \qquad (15.17)$$

(Note that $R_{33} = CT_{33}$ gives rise to the same equation and is redundant.) Using (15.16), we can write this relation as

$$e^{\nu}[1 + r\nu'] = 1 + \frac{C\epsilon^2}{2c^2r^2} \qquad (15.18)$$

Observe that differentiating this gives rise to (15.14*a*), so (15.14*a*), (15.14*b*), and (15.18) are consistent equations. This differential equation is easily solved for ν by noting that the left side is the derivative of re^{ν}. Thus

$$(re^{\nu})' = 1 + \frac{C\epsilon^2}{2c^2r^2} \qquad (15.19)$$

which integrates immediately to give

$$e^{\nu} = 1 - \frac{2m}{r} - \frac{C\epsilon^2}{2c^2r^2} \qquad (15.20)$$

The constant of integration $2m$ is the same as in the ordinary Schwarzschild solution, namely $2\kappa M/c^2$, since we have to ensure that in the case $\epsilon = 0$ our new solution will coincide with the original Schwarzschild form.

Let us now collect our results. The line element, from (15.16) and (15.20), is

$$ds^2 = \left(1 - \frac{2m}{r} - \frac{C\epsilon^2}{2c^2r^2}\right) c^2\,dt^2 - \left(1 - \frac{2m}{r} - \frac{C\epsilon^2}{2c^2r^2}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2) \qquad (15.21)$$

where $m = \kappa M/c^2$ and $C = -8\pi\kappa/c^2$. The radial electric field is, from (15.11) and (15.16),

$$E = \frac{\epsilon}{r^2} \qquad (15.22)$$

The above results were first obtained by Nordström (1918) and Reissner (1916). It is interesting to note that, since C is negative, the function

$$e^{\nu} = \left(1 - \frac{2m}{r} - \frac{C\epsilon^2}{2c^2r^2}\right)$$

takes the form shown in Fig. 15.1. In particular, if ϵ^2 is sufficiently large $\left(\text{more precisely, if } \left(\frac{\epsilon}{M}\right)^2 \frac{1}{\kappa} > \frac{1}{4\pi}\right)$, no singular sphere exists, unlike the case of the ordinary Schwarzschild solution, which possesses a singular sphere at $r = 2m$. For a proton, one obtains $(\epsilon/M)^2/\kappa \cong 10^{36}$, so a

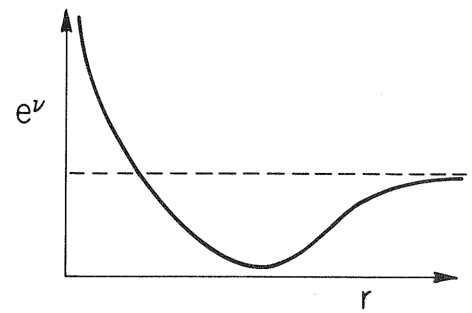


Fig. 15.1

proton has no singular sphere. It is also remarkable that the effect of the electromagnetic charge on the metric dies off faster at infinity than the effect of the gravitational mass (see Exercise 15.1).

15.2 Weyl's Generalization of Riemannian Geometry

The general theory of relativity succeeded in geometrizing the phenomenon of gravitation by connecting it with the metric of the Riemann space considered. The potential of the gravitational force which occurs in the Newtonian theory was replaced by the metric potentials $g_{\mu\nu}$, the components of the metric tensor. If we wish to obtain an analogous theory for electromagnetic phenomena, we have to establish corresponding relations between the electromagnetic potentials and the metric tensor. However, the components $g_{\mu\nu}$ of the metric tensor are already sufficiently determined by the Einstein field equations, and there seems to be no room to imbed also the entire theory of the electromagnetic field into the same differential geometry.

In 1918 Weyl proposed a generalization of differential geometry which allows a greater freedom in the choice of a metric tensor, and this freedom appeared just large enough to imbed the entire electromagnetic formalism into the new geometry. While the success of this theory was rather limited from the physical point of view, it showed an interesting possibility of a generalized differential geometry which contains a suggestive formalism and may still have the germs of a future fruitful theory. We shall therefore give a brief outline of Weyl's ideas (Weyl, 1918).

We start again with the idea of an affine vector transplantation as in Sec. 2.1. That is, we ask for a law of vector transplantation between different points of the manifold which appears locally and in a properly chosen local coordinate system as a transplantation of unchanged vector components. As was shown in Sec. 2.1, such a transplantation law appears in an arbitrary coordinate system in the differential form

$$d\xi^\alpha = \Gamma^\alpha{}_{\beta\gamma}\, dx^\beta\, \xi^\gamma \tag{15.23}$$

where the $\Gamma^\alpha{}_{\beta\gamma}$ are the symmetric connections of the manifold, ξ^α are the components of the vector considered, and dx^β is the local displacement vector.

Next we assume again the existence of a symmetric tensor field $g_{\mu\nu}$ which serves as the metric tensor. Thus, at every point of the manifold we can determine the length l of the vector ξ^α by means of the formula

$$l^2 = \|\xi\|^2 = g_{\alpha\beta}\xi^\alpha\xi^\beta \tag{15.24}$$

Analogously, we can calculate the scalar product of two vectors, say, ξ^α and η^α, attached to the point considered by

$$\xi^\alpha\eta_\alpha = g_{\alpha\beta}\xi^\alpha\eta^\beta \tag{15.25}$$

In Sec. 2.2 we introduced at this stage the requirement that the length of a vector and the scalar product of two vectors should remain unchanged under the transplantation law (15.23). This postulate led us to the determination of the connections as the Christoffel symbols of the metric tensor and resulted in the classical Riemannian differential geometry on the manifold.

It is at this stage that the Weyl modification of the differential geometry sets in. We do not demand conservation of length and scalar products under affine transplantation. If we interpret the vector ξ^α as a physical measuring rod with prescribed orientation and assume that it changes under transport from point to point in the manifold according to the law (15.23), our relaxation of this requirement means that we allow the rod to change length under bodily displacement. This dropping of the restrictive demand of length preservation opens up a greater freedom in the choice of a differential geometry, and it is here that Weyl succeeds in bringing in geometric quantities which he will identify with electromagnetic potentials.

Of course, we have to make some assumptions regarding the length of transplanted vectors if we want to specialize the rather structureless theory of affine connections. It is natural to assume in analogy to (15.23) that the increment in length is proportional to the length itself and a linear homogeneous function of the displacement vector dx^α. Hence, we set up

$$dl = (\varphi_\beta\, dx^\beta)l \tag{15.26}$$

Here the covariant vector φ_β plays a role analogous to that of the connection $\Gamma^\alpha{}_{\beta\gamma}$. Combining (15.26) with (15.23) and (15.24), we obtain

$$\begin{aligned} dl^2 &= 2l^2(\varphi_\beta\, dx^\beta) = d(g_{\alpha\beta}\xi^\alpha\xi^\beta) \\ &= g_{\alpha\beta|\gamma}\xi^\alpha\xi^\beta\, dx^\gamma + g_{\alpha\beta}\Gamma^\alpha{}_{\rho\gamma}\xi^\rho\xi^\beta\, dx^\gamma + g_{\alpha\beta}\Gamma^\beta{}_{\rho\gamma}\xi^\alpha\xi^\rho\, dx^\gamma \end{aligned} \tag{15.27}$$

Renaming the various summation indices, rearranging terms, and using (15.24) again, we can bring (15.27) into the form

$$[g_{\alpha\beta|\gamma} + g_{\sigma\beta}\Gamma^\sigma{}_{\alpha\gamma} + g_{\alpha\sigma}\Gamma^\sigma{}_{\beta\gamma}]\xi^\alpha\xi^\beta\, dx^\gamma = 2g_{\alpha\beta}\varphi_\gamma\xi^\alpha\xi^\beta\, dx^\gamma \tag{15.28}$$

Since (15.28) must hold for arbitrary choice of ξ^α and dx^γ, we conclude in the usual manner that

$$(g_{\alpha\beta|\gamma} - 2g_{\alpha\beta}\varphi_\gamma) + g_{\sigma\beta}\Gamma^\sigma{}_{\alpha\gamma} + g_{\sigma\alpha}\Gamma^\sigma{}_{\beta\gamma} = 0 \tag{15.29}$$

This is the same system of linear equations for the connections $\Gamma^\alpha{}_{\beta\gamma}$ as in Sec. 2.2; only the inhomogeneous term $g_{\alpha\beta|\gamma}$ has now to be replaced by $g_{\alpha\beta|\gamma} - 2g_{\alpha\beta}\varphi_\gamma$. Hence, the same linear algebra as in Sec. 2.2 leads to the equations

$$\Gamma^\alpha{}_{\beta\gamma} = -\begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix} + g^{\sigma\alpha}[g_{\sigma\beta}\varphi_\gamma + g_{\sigma\gamma}\varphi_\beta - g_{\beta\gamma}\varphi_\sigma] \tag{15.30}$$

Here $\begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix}$ is the usual Christoffel symbol of the second kind. Thus we may arbitrarily prescribe the metric-tensor field $g_{\alpha\beta}$ and the covariant vector field φ_α and determine by (15.30) the field of connections $\Gamma^\alpha{}_{\beta\gamma}$

which admits under the affine transplantation law (15.23) the length transplantation rule (15.26). Clearly, the differential geometry obtained is a generalization of the Riemann geometry discussed until now. If we select the vector field $\varphi_\beta \equiv 0$, the Weyl geometry reduces to the classical Riemann geometry.

Let us point out that the mathematical theory of transplantation of vectors and length is, according to (15.23) and (15.26), very useful even in the case of classical Riemannian differential geometry. It allows us a greater flexibility in our choice of the metric tensor $g_{\alpha\beta}$. Indeed, let $f(x^\lambda)$ be a scalar field on the manifold, and let us introduce the new metric tensor

$$(15.31) \qquad \hat{g}_{\alpha\beta} = f(x^\lambda) g_{\alpha\beta}$$

In the new metric, a vector ξ^α would have the length l given by

$$(15.32) \qquad \hat{l}^2 = \hat{g}_{\alpha\beta}\xi^\alpha\xi^\beta = f(x^\lambda) g_{\alpha\beta}\xi^\alpha\xi^\beta = f(x^\lambda) l^2$$

where l is the length of the same vector as measured in the classical way by means of the original metric tensor $g_{\alpha\beta}$.

In the original metric with the tensor $g_{\alpha\beta}$ we have

$$(15.33) \qquad \varphi_\alpha \equiv 0 \qquad \Gamma^\alpha{}_{\beta\gamma} = -\left\{ {\alpha \atop \beta\ \gamma} \right\}$$

and the length l of a vector is unchanged under parallel displacement. However, the same displacement law in the metric $\hat{g}_{\alpha\beta}$ leads to the relation

$$(15.34) \qquad \frac{d\hat{l}}{\hat{l}} = \tfrac{1}{2}(\log f)_{|\lambda}\, dx^\lambda$$

as can be seen from (13.32). Thus $\frac{1}{2}(\log f)_{|\lambda}$ plays the role of φ_λ in (15.26); it then follows that the ordinary connections $-\left\{ {\alpha \atop \beta\ \gamma} \right\}$ constructed from $g_{\alpha\beta}$ are equal to the more general connections $\hat{\Gamma}^\alpha{}_{\beta\gamma}$ constructed according to (15.30) from $\hat{g}_{\alpha\beta}$ and $\varphi_\lambda = \frac{1}{2}(\log f)_{|\lambda}$:

$$(15.35) \qquad \hat{\Gamma}^\alpha{}_{\beta\gamma} = \Gamma^\alpha{}_{\beta\gamma}$$

as can also be seen by direct computation from (15.30) and (15.31). But in the new metric we have a length transplantation with

$$(15.36) \qquad \hat{\varphi}_\lambda = \frac{1}{2}\frac{\partial}{\partial x^\lambda} \log f$$

Thus we can deal with a Riemannian manifold in a more flexible manner by use of a more general metric tensor $\hat{g}_{\alpha\beta}$ if we readjust the length measurement by the length transplantation law (15.26) with the corresponding vector field (15.36). We may interpret the change of metric from $g_{\alpha\beta}$ to $\hat{g}_{\alpha\beta}$ by (15.31) as a change of scale for the length at every point of the Riemann manifold by the variable gauge factor $f(x^\lambda)$. The substitution (15.31) is therefore called a gauge transformation, and $\varphi_\alpha(x^\lambda)$ is called the gauge vector field.

Our generalized differential geometry separates neatly the problem of measurement of angles from that of measurement of length. Indeed, the angle between the two vectors ξ^α and η^α at a given point of the manifold is measured by the ratio

$$(15.37) \qquad \frac{\xi^\alpha\eta_\alpha}{\|\xi\|\ \|\eta\|} = \frac{g_{\alpha\beta}\xi^\alpha\eta^\beta}{[(g_{\alpha\beta}\xi^\alpha\xi^\beta)(g_{\alpha\beta}\eta^\alpha\eta^\beta)]^{1/2}}$$

This ratio does not change under the gauge transformation (15.31). The gauge transformation is therefore a conformal (i.e., angle-preserving) change of the metric. On the other hand, the length of vectors will change under (15.31) according to (15.32). Thus the metric tensor $g_{\alpha\beta}$ determines angles, while one needs also the gauge vector φ_α to measure length.

We return now from the case of a Riemannian manifold to the general case of a Weyl geometry which is characterized by an arbitrary symmetric tensor field $g_{\alpha\beta}$ and an arbitrary gauge vector field φ_α. The connections $\Gamma^\alpha{}_{\beta\gamma}$ are then determined by Eqs. (15.30). The same argument as before shows that we may replace the geometric quantities by use of a scalar field f as follows:

$$(15.38) \qquad \hat{g}_{\alpha\beta} = f(x^\lambda) g_{\alpha\beta} \qquad \hat{\varphi}_\alpha = \varphi_\alpha + \tfrac{1}{2}(\log f)_{|\alpha} \qquad \hat{\Gamma}^\alpha{}_{\beta\gamma} = \Gamma^\alpha{}_{\beta\gamma}$$

without changing the intrinsic geometric properties of vector fields. That is, in the new metric, vectors will have the same law of affine transplantation and the angle between different vectors at the same point of the manifold will be preserved, but the local lengths of a vector will be changed according to

$$(15.39) \qquad \hat{l}^2 = f(x^\lambda) l^2$$

Thus the general Weyl geometry admits also a conformal gauge transformation, which is, of course, in this case of much greater significance. Quantities and relations that do not change under gauge transformations are called gauge invariants.

Given an arbitrary metric field consisting of the metric tensor $g_{\alpha\beta}$ and the gauge vector φ_α, the question arises whether or not this field may be transformed into a metric field of ordinary Riemannian geometry by means of a gauge transformation (15.38). That is, we have to decide if, by a proper choice of the gauge factor $f(x^\lambda)$, we can achieve $\hat{\varphi}_\alpha \equiv 0$. Clearly, we can reduce $\hat{\varphi}_\alpha$ to the zero vector if and only if φ_α is a gradient vector field, that is, if

$$\varphi_{\alpha|\beta} - \varphi_{\beta|\alpha} \equiv 0 \tag{15.40}$$

Condition (15.40) may also be expressed in the following suggestive manner: The necessary and sufficient condition that a Weyl geometry may be reduced to a Riemannian geometry is that a vector keep its original length after transplantation along an arbitrary closed trajectory. Indeed, the condition of such a length preservation is, by (15.26),

$$\oint_C \frac{dl}{l} = \oint_C \varphi_\alpha \, dx^\alpha = 0 \tag{15.41}$$

and it is well known that (15.40) is the necessary and sufficient condition for the integrability requirement (15.41) in simply connected regions.

We are thus led to study the tensor field

$$F_{\alpha\beta} = \varphi_{\alpha|\beta} - \varphi_{\beta|\alpha} \tag{15.42}$$

It stands in close analogy to the Riemann curvature tensor field $R_{\alpha\beta\gamma\delta}$. Just as the vanishing of the latter tensor field is the necessary and sufficient condition that a vector return into itself after transplantation along a closed trajectory, so the vanishing of the tensor field $F_{\alpha\beta}$ is the necessary and sufficient condition that the length of a vector be preserved under such transplantation. As we showed in Chap. 5, the vanishing of the Riemann tensor guarantees a choice of coordinates in which the geometry becomes pseudo-Euclidean. Likewise, the vanishing of $F_{\alpha\beta}$ guarantees a choice of metric in which the Weyl geometry becomes Riemannian.

The Riemann curvature tensor has specific symmetry properties and satisfies the Bianchi differential relations. Similarly, the new tensor field is antisymmetric:

$$F_{\alpha\beta} = -F_{\beta\alpha} \tag{15.43}$$

and satisfies the differential equations

$$\{F_{\alpha\beta|\gamma}\} = 0 \tag{15.44}$$

Clearly, $F_{\alpha\beta}$ is the intrinsic geometric quantity of the Weyl geometry.

Indeed, under a gauge transformation (15.38) the gauge vector field φ_α will change, but its curl vector field $F_{\alpha\beta}$ will be unchanged.

We have now introduced the most important concepts of Weyl's generalized differential geometry and can repeat most of the general ideas of the Riemannian geometry. This is possible since we still have the concept of vector transplantation. Thus we define again a geodesic line as a curve whose tangent vector is carried along that curve by the law of vector transplantation, i.e., remains parallel to itself. We had in Riemannian geometry a second definition of a geodesic, namely, as a line between two points, which made the curve length stationary under variation. This latter definition cannot be used in Weyl's geometry since the curve length is not a gauge-invariant quantity. On the other hand, the concept of a null geodesic is obviously gauge-invariant. This is an important fact in view of the central role of null geodesics in general relativity theory.

Likewise, the concept of covariant differentiation depends only on the concept of vector transplantation. Indeed, we measure the change of a vector component between two nearby points by comparing the actual component after displacement with the value of the same component which we should have obtained under affine transplantation. Thus we may define

$$\xi^\alpha{}_{\|\beta} = \xi^\alpha{}_{|\beta} - \Gamma^\alpha{}_{\beta\gamma}\xi^\gamma \tag{15.45}$$

In Riemannian differential geometry, the curvature tensor $R^\alpha{}_{\beta\gamma\delta}$ was introduced through the law of interchange in the order of covariant differentiations:

$$\xi^\alpha{}_{\|\beta\|\gamma} - \xi^\alpha{}_{\|\gamma\|\beta} = R^\alpha{}_{\eta\beta\gamma}\xi^\eta \tag{15.46}$$

Hence, we can now express the curvature tensor in terms of the connections $\Gamma^\alpha{}_{\beta\gamma}$ in precisely the same manner as we did in Chap. 5, by means of the special connections, the Christoffel symbols,

$$R^\alpha{}_{\beta\gamma\delta} = -\Gamma^\alpha{}_{\beta\gamma|\delta} + \Gamma^\alpha{}_{\beta\delta|\gamma} + \Gamma^\alpha{}_{\tau\delta}\Gamma^\tau{}_{\beta\gamma} - \Gamma^\alpha{}_{\tau\gamma}\Gamma^\tau{}_{\beta\delta} \tag{15.47}$$

Using the expression (15.30) for the connections in terms of the metric tensor and the gauge vector, we can thus express the curvature tensor in terms of these basic metric fields.

While the complete expression for the curvature tensor is rather involved, it is relatively easy to give a closed formula for the curvature scalar R defined by double contraction:

$$R_{\beta\delta} = R^\alpha{}_{\beta\alpha\delta} \qquad R = g^{\beta\delta}R_{\beta\delta} \tag{15.48}$$

Indeed, since the form of the scalar is independent of the coordinate system used, we may compute it in a coordinate system which is geodesic at the point considered. We denote the Christoffel symbols in the corresponding Riemann space by

$$\dot{\Gamma}^{\alpha}{}_{\beta\gamma} = -\left\{\begin{matrix}\alpha \\ \beta\ \gamma\end{matrix}\right\} \tag{15.49}$$

and denote the Riemannian curvature tensor (built by a formula analogous to (15.47) but based on the $\dot{\Gamma}^{\alpha}{}_{\beta\gamma}$ instead of the actual connections $\Gamma^{\alpha}{}_{\beta\gamma}$) by $\dot{R}^{\alpha}{}_{\beta\gamma\delta}$; analogously, we define the contracted tensors $\dot{R}_{\beta\delta}$ and $\dot{R}$. By our assumption all $\dot{\Gamma}^{\alpha}{}_{\beta\gamma}$ and all $g^{\alpha\beta}{}_{|\gamma}$ vanish at the point where we wish to determine the curvature scalar R. Hence, we have, by (15.30),

$$\Gamma^{\alpha}{}_{\beta\gamma} = \varphi_{\gamma}\delta^{\alpha}{}_{\beta} + \varphi_{\beta}\delta^{\alpha}{}_{\gamma} - g_{\beta\gamma}\varphi^{\alpha} \tag{15.50}$$

We derive from (15.47) and (15.48) the equation

$$R = -g^{\beta\delta}\Gamma^{\alpha}{}_{\beta\alpha|\delta} + (g^{\beta\delta}\Gamma^{\alpha}{}_{\beta\delta})_{|\alpha} + g^{\beta\delta}\Gamma^{\alpha}{}_{\tau\delta}\Gamma^{\tau}{}_{\beta\alpha} - \Gamma^{\alpha}{}_{\tau\alpha}g^{\beta\delta}\Gamma^{\tau}{}_{\beta\delta} \tag{15.51}$$

An easy calculation based on (15.30) yields the identities

$$g^{\beta\delta}\Gamma^{\alpha}{}_{\beta\delta} = g^{\beta\delta}\dot{\Gamma}^{\alpha}{}_{\beta\delta} - (n-2)\varphi^{\alpha} \tag{15.52}$$

and

$$\Gamma^{\alpha}{}_{\beta\alpha} = \dot{\Gamma}^{\alpha}{}_{\beta\alpha} + n\varphi_{\beta} \tag{15.53}$$

valid at all points; here n is the dimension of the space considered and $n = 4$ in the case of general relativity.

We also find at the point considered, by virtue of (15.50),

$$g^{\beta\delta}\Gamma^{\alpha}{}_{\tau\delta}\Gamma^{\tau}{}_{\beta\alpha} = -(n-2)(\varphi_{\alpha}\varphi^{\alpha}) \tag{15.54}$$

Inserting all these terms into (15.51), we arrive at

$$R = \dot{R} + (n-1)(n-2)(\varphi_{\alpha}\varphi^{\alpha}) - 2(n-1)\varphi^{\alpha}{}_{|\alpha} \tag{15.55}$$

We can bring the right-hand side into covariant form using (3.12):

$$R = \dot{R} + (n-1)(n-2)(\varphi_{\alpha}\varphi^{\alpha}) - 2(n-1)\frac{1}{\sqrt{-g}}(\sqrt{-g}\,\varphi^{\alpha})_{|\alpha} \tag{15.56}$$

Indeed, the right-hand side is formally invariant under any change of coordinates. Since both sides of (15.56) are scalars, the identity (15.56) is valid in every coordinate system. We have thus expressed the curvature scalar R in an elegant way in terms of the former Riemann scalar $\dot{R}$ and of the gauge vector φ_{α}. Formula (15.56) is due to Weyl and will play a role in the theory of electromagnetism to be developed in the next section. It will, of course, be needed only for the dimension $n = 4$.

There is one important point to be observed in the tensor algebra of the Weyl geometry. We may consider a vector field ξ^{α} given independently of the metric used. However, if we form from this contravariant field the covariant field

$$\xi_{\alpha} = g_{\alpha\beta}\xi^{\beta} \tag{15.57}$$

then this new vector field will depend upon the metric, and under a gauge transformation (15.38) we shall have

$$\hat{\xi}_{\alpha} = f(x^{\lambda})\xi_{\alpha} \tag{15.58}$$

Thus the covariant form of a gauge-invariant contravariant vector becomes gauge-dependent. Weyl introduced the concept of the "weight" of a tensor relative to gauge transformations. We say that a tensor is of weight n if it is multiplied under a gauge transformation (15.38) by the factor $f(x^{\lambda})^n$:

$$\hat{T}^{\alpha\cdots}{}_{\beta\cdots} = f(x^{\lambda})^n T^{\alpha\cdots}{}_{\beta\cdots} \tag{15.59}$$

Clearly, the usual manipulation of indices in vector algebra will introduce tensors of various weights. Observe, however, that the gauge vector φ_{α} plays here a singular role since its gauge transformation is described by the particular law (15.38). Therefore we cannot ascribe a weight to the gauge vector.

Similarly, we can assign weights to tensor densities. Indeed, the density factor $\sqrt{-g}$ transforms in four-dimensional space according to

$$\sqrt{-\hat{g}} = f^2\sqrt{-g} \tag{15.60}$$

and has the weight 2.

We may consider, for example, the gauge-invariant antisymmetric tensor (15.42). Its doubly contravariant form

$$F^{\alpha\beta} = g^{\alpha\mu}g^{\beta\nu}F_{\mu\nu} \tag{15.61}$$

is easily seen to be of weight $n = -2$. But the corresponding density

$$\mathfrak{F}^{\alpha\beta} = F^{\alpha\beta}\sqrt{-g} \tag{15.62}$$

is, by virtue of (15.60), of weight zero, and hence gauge-invariant. Similarly, the scalar density $F_{\alpha\beta}F^{\alpha\beta}\sqrt{-g}$ which occurred in the variational principle in Sec. 11.4 is gauge-invariant.

Finally, we observe that the curvature tensor (15.47) is gauge-invariant, that the same is true for its contracted form $R_{\beta\delta}$, but that the curvature scalar R has the weight -1. Hence, the scalar density $R^2\sqrt{-g}$ is gauge-invariant. This fact will play a role in the considerations of the next section, where we shall search for gauge-invariant Lagrangians in possible variational principles. But we recognize already at this stage that the concept of gauge invariance singles out particularly important expressions among tensors and tensor densities.

15.3 Weyl's Theory of Electromagnetism

In the preceding sections we discussed the generalized differential geometry of Weyl as a logically possible and formally elegant mathematical theory. In this section we shall now indicate the physical interpretation of this geometry and its connection with the electromagnetic and gravitational fields of the space considered.

We wish to geometrize electromagnetism and gravitation at the same time. We therefore have to express these fields in terms of the metric of physical space. Let us start with the problem of the electromagnetic field. Equations (15.42) and (15.44) are so similar to Maxwell's equations relating the vector potential with the electromagnetic tensor that it is natural to interpret φ_α and $F_{\alpha\beta}$ in just this manner. Thus, according to Weyl, the electromagnetic field with the vector potential φ_α induces a geometry with a gauge vector proportional to φ_α, or conversely, the dynamical effect of such a geometry is the same as that of the electromagnetic field in classical interpretation. The set of Maxwell's equations

$$\{F_{\alpha\beta|\gamma}\} = 0 \tag{15.63}$$

is automatically fulfilled, while the complementary set

$$\mathfrak{F}^{\alpha\beta}{}_{|\beta} = \mathfrak{s}^\alpha \tag{15.64}$$

is gauge-invariant, in view of our remarks at the end of the last section. We thus always have to interpret the vector $\mathfrak{s}^\alpha$ of current density as gauge-independent.

We pointed out in Sec. 4.1 that Maxwell's equations are unchanged if we introduce an arbitrary change of scale in the metric. We may now state this result in the following form: Maxwell's equations are gauge-invariant. This fact thus obtains added significance in the Weyl theory. It is a natural consequence of our geometric interpretation of the electromagnetic field, and no longer only a mathematical accident. It should also be recalled at this point that the other massless fields which one can describe (the neutrino field and massless fields of arbitrary spin) are also represented by gauge-invariant equations (Penrose, 1964).

If we now wish to discuss the interaction between the electromagnetic and the gravitational field, we shall have to set up field equations between the 14 field quantities φ_α and $g_{\alpha\beta}$. We shall do so by setting up an action integral

$$I = \int W\sqrt{-g}\,d^4x \tag{15.65}$$

which is based on the field quantities φ_α and $g_{\alpha\beta}$ and deriving the field equations as the Euler-Lagrange equations of the variational problem

$$\delta I = 0 \tag{15.66}$$

where the field variables φ_α and $g_{\alpha\beta}$ are varied independently. The integrand

$$\mathfrak{W} = W\sqrt{-g} \tag{15.67}$$

is assumed to be a scalar density of weight zero since we wish to obtain gauge-invariant field equations. Without specifying yet the explicit structure of $\mathfrak{W}$, we introduce its functional derivatives with respect to the field variables by the identity

$$\delta\int \mathfrak{W}\,d^4x = \int(\mathfrak{W}^\alpha\,\delta\varphi_\alpha + \mathfrak{W}^{\alpha\beta}\,\delta g_{\alpha\beta})\,d^4x \tag{15.68}$$

where we assume that all variations of the field variables vanish at the boundary. The quantities $\mathfrak{W}^\alpha$ and $\mathfrak{W}^{\alpha\beta}$ can easily be calculated by the Euler-Lagrange equations once the structure of the integrand $\mathfrak{W}$ has been specified.

The field equations will then take the form

$$\mathfrak{W}^\alpha = 0 \qquad \mathfrak{W}^{\alpha\beta} = 0 \tag{15.69}$$

as follows from (15.66) and (15.68). These 14 equations are not independent of each other. Indeed, the formal structure of $\mathfrak{W}$ guarantees that the integral I will not change if we make an arbitrary infinitesimal change of coordinates or an infinitesimal gauge transformation. Thus the fact that $\mathfrak{W}$ is a scalar density of weight zero establishes a priori various relations between the functional derivatives. For example, let us make a gauge transformation (15.38) with the infinitesimal change of scale

$$f(x^\lambda) = 1 + \epsilon\pi(x^\lambda) \tag{15.70}$$

According to (15.38) this implies the variations of the field variables

$$\delta g_{\alpha\beta} = \epsilon\pi g_{\alpha\beta} \qquad \delta\varphi_\alpha = \tfrac{1}{2}\epsilon\pi_{|\alpha} \tag{15.71}$$

We have to assume that $\pi(x^\lambda)$ vanishes on the boundary of our domain of integration if we wish to apply the identity (15.68). But otherwise $\pi(x^\lambda)$ is quite arbitrary. Hence, we conclude from the gauge invariance of $\mathfrak{W}$ that

$$\int[\mathfrak{W}^{\alpha\beta}g_{\alpha\beta} - \tfrac{1}{2}\mathfrak{W}^\alpha{}_{|\alpha}]\pi\, d^4x = 0 \tag{15.72}$$

for every admissible variation, and consequently that

$$\mathfrak{W}^\alpha{}_{|\alpha} = 2\mathfrak{W}^\alpha{}_\alpha \tag{15.73}$$

This identity is a formal consequence of the gauge invariance of $\mathfrak{W}$ and not a physical law, as are most of the remaining field equations (15.69).

Similarly, we can derive four additional formal identities by considering infinitesimal changes of variables,

$$\bar{x}^\alpha = x^\alpha + \epsilon\xi^\alpha \tag{15.74}$$

which depend on the four arbitrary functions ξ^α. Thus the field equations (15.69) give only nine independent and physically significant statements about the actual field relations.

After these general considerations we now have the problem of selecting a specific Lagrangian density $\mathfrak{W}$ of weight zero which will lead to physically acceptable field equations. We shall assume that $\mathfrak{W}$ is built of the components of the metric tensor $g_{\alpha\beta}$ and their first and second derivatives, and furthermore that it contains the components φ_α of the gauge vector and their first derivatives. Then it can be shown that the only rational functions of these terms which are scalar densities of weight zero are the following expressions:

$$F_{\alpha\beta}F^{\alpha\beta}\sqrt{-g}, \quad R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta}\sqrt{-g}, \quad R_{\alpha\beta}R^{\alpha\beta}\sqrt{-g}, \quad R^2\sqrt{-g} \tag{15.75}$$

Observe that we derived in Sec. 11.4 a variational principle with the Lagrangian density

$$(R + \tfrac{1}{2}CF_{\alpha\beta}F^{\alpha\beta})\sqrt{-g} \tag{15.76}$$

which led indeed to Einstein's and Maxwell's field equations. However, in this case the curvature scalar R was the Riemannian scalar and was independent of the vector potential φ_α of the electromagnetic field. The Lagrangian (15.76) is unacceptable in Weyl's theory since the term $R\sqrt{-g}$ is of weight 1 and not gauge-invariant. The closest admissible analogy to the Lagrangian density (15.76) in Weyl's theory would be the expression

$$\mathfrak{W} = (R^2 + AF_{\alpha\beta}F^{\alpha\beta})\sqrt{-g} \tag{15.77}$$

The field equations can then be expressed by the variational condition

$$\begin{aligned}\delta\int\mathfrak{W}\, d^4x &\\ = \int[2R\,\delta R\sqrt{-g} &+ R^2\delta\sqrt{-g} + A\,\delta(F_{\alpha\beta}F^{\alpha\beta}\sqrt{-g})]\, d^4x\end{aligned} \tag{15.78}$$

where the $g_{\alpha\beta}$ and φ_α are varied independently and arbitrarily except for the requirement that they vanish on the boundary of the domain of integration.

We now can considerably simplify the expression (15.78) by using the fact that the scalar R has the weight -1. Indeed, we can introduce a local scale of length into the metric field considered such that

$$R = \lambda \tag{15.79}$$

is constant. Of course, this particular gauge will be destroyed after the arbitrary variation of the metric field and $\delta R \neq 0$ in general. The constant λ is a measure of the curvature of the space and corresponds roughly to the cosmological constant considered in Chap. 13.

We now can bring (15.78) into the elegant form

$$\delta\int\left[R + \frac{1}{2\lambda}AF_{\alpha\beta}F^{\alpha\beta} - \tfrac{1}{2}\lambda\right]\sqrt{-g}\, d^4x = 0 \tag{15.80}$$

whose Lagrangian density is now very similar to the original form (15.76) since it depends linearly on R.

Finally, we turn to the expression (15.56) of R in terms of the Christoffel symbols. We put $n = 4$ and observe that

$$\int R\sqrt{-g}\, d^4x = \int[\mathring{R} + 6(\varphi_\alpha\varphi^\alpha)]\sqrt{-g}\, d^4x - 6\int(\varphi^\alpha\sqrt{-g})_{|\alpha}\, d^4x \tag{15.81}$$

The second right-hand integral depends only upon the values of the integrand on the boundary and does not change under the variations considered. Thus we obtain, from (15.80) and (15.81),

$$\delta \int \left[\dot{R} + \frac{1}{2}\frac{A}{\lambda} F_{\alpha\beta}F^{\alpha\beta} + 6(\varphi_\alpha \varphi^\alpha) - \tfrac{1}{2}\lambda \right] \sqrt{-g}\, d^4x = 0 \tag{15.82}$$

For convenience we define

$$\varphi_\alpha = \sqrt{\lambda}\, \tilde{\varphi}_\alpha \qquad F_{\alpha\beta} = \sqrt{\lambda}\, \tilde{F}_{\alpha\beta} \tag{15.83}$$

where $\tilde{\varphi}_\alpha$ and $\tilde{F}_{\alpha\beta}$ are the components of the vector potential and the electromagnetic field measured in usual units. We then have

$$\delta \int [R + \tfrac{1}{2} A \tilde{F}_{\alpha\beta}\tilde{F}^{\alpha\beta} - \lambda(\tfrac{1}{2} - 6\tilde{\varphi}_\alpha \tilde{\varphi}^\alpha)] \sqrt{-g}\, d^4x = 0 \tag{15.84}$$

The first two terms in the integral now correspond precisely to the terms in the Lagrangian (15.76) of the classical theory, while the correction term is in general small because of the smallness of the cosmological term λ. In most astronomical problems the effect of the correction terms is negligible, and only in the theory of the elementary particles might these terms be of importance.

Let us consider the effect of varying the components φ_α of the gauge vector. The same calculations as were used in Sec. 11.4 lead to the equations

$$A \tilde{\mathfrak{F}}^{\alpha\beta}{}_{|\beta} = 6\lambda \tilde{\varphi}^\alpha \sqrt{-g} \tag{15.85}$$

We can identify these equations with the set of Maxwell's equations (4.66) if we define the vector of current density

$$\tilde{\mathfrak{J}}^\alpha = \frac{6\lambda}{A} \tilde{\varphi}^\alpha \sqrt{-g} \tag{15.86}$$

From the Maxwell equations themselves follows the law of conservation of charge:

$$\tilde{\mathfrak{J}}^\alpha{}_{|\alpha} = 0 \tag{15.87}$$

Let us observe that the expression (15.86) for the current density vector is valid only in a metric in which the scalar R is constant. In order to put (15.86) into gauge-invariant form we observe that, by (15.38) and the fact that R has the weight -1, it follows that the

expression

$$\varphi_\alpha + \tfrac{1}{2}(\log R)_{|\alpha} \tag{15.88}$$

is gauge-invariant. Thus we may put (15.86) into the form

$$\tilde{\mathfrak{J}}^\alpha = \frac{6}{A\sqrt{\lambda}} (R\varphi_\beta + \tfrac{1}{2}R_{|\beta}) g^{\alpha\beta} \sqrt{-g} \tag{15.89}$$

which is valid in every admissible metric. This formula shows the interrelation between the metric quantities and the sources of the electromagnetic field.

While the mathematical formalism of Weyl's theory of electromagnetism has a high degree of consistency and elegance, it has led to no prediction of new physical phenomena which could be observed and might serve as confirmations of the theory. On the contrary, in an appendix to Weyl's exposition of his unified field theory, Einstein raised some very serious objections to it on empirical physical grounds. Let us consider the case of a static, radially symmetric gravitational field in which a nonzero electrostatic field is present, which is also assumed as time-independent and radially symmetric. In this case the gauge vector φ_α will have only one nonvanishing component, namely, φ_0. This function will depend only upon the distance from the center of symmetry. Next we put a clock at a given fixed point of this field. It measures time by means of a periodic process, which has the duration τ_0 in the time marker x^0. The physical time coincides for this resting clock with the quantity $(1/c)l$, and according to (15.26) we shall have, after the time x^0 has passed,

$$l = l_0 \exp\left(\int_0^{x^0} \varphi_0\, dx^0\right) = l_0 \exp(\varphi_0 x^0) \tag{15.90}$$

We may choose as l_0 the period τ_0 of the clock, assuming that, at the moment $x^0 = 0$, the marker interval coincides with the physical time interval. However, after the time x^0 has passed, the physical measure of a period of the clock will be given by

$$\tau = \tau_0 \exp(\varphi_0 x^0) \tag{15.91}$$

In particular, if we had two identical clocks at two points of the field with different values of φ_0, these clocks would differ more and more in frequency as time goes on. We might consider this effect on atomic clocks and should then expect that the frequency of the various spectral lines should depend on the location and past histories of the atoms. But it is a well-known fact that the spectral lines are very sharp and well defined; whence Einstein concluded that Weyl's theory is in contradiction to experience. The strength of Einstein's objection seems not as powerful now as at the time when it was raised, since we know that classical physics does not describe atomic phenomena without certain quantum-theoretical modifications. However, it seems indeed strange that two identical physical systems at the same point in space-time should be different because of different past histories.

An interesting consideration regarding Weyl's theory of length transport and gauge invariance is due to London (1927). He considers the motion of an electron in the field of a proton and applies to it the gauge concept. We obviously have

$$\varphi_0 = \frac{\alpha}{r} \qquad \varphi_i \equiv 0 \tag{15.92}$$

where r is the distance from the proton, and α is a dimensionless constant of proportionality which connects the $1/r$ electrostatic potential of the proton with the geometric gauge term φ_0. Let us consider circular motion for the sake of simplicity. We have the equality between centrifugal and electrostatic forces,

$$\frac{mv^2}{r} = \frac{e^2}{r^2} \tag{15.93}$$

from which we can compute the time for describing one orbit:

$$T = \frac{2\pi r}{v} \qquad v = \frac{e}{\sqrt{mr}} \tag{15.94}$$

During this orbit the scale of length has changed according to (15.90). London raises the question whether or not it is possible that this change vanishes for certain orbits and for proper choice of the numerical factor α. We are led to the condition

$$\exp(\varphi_0 cT) = 1 \qquad \varphi_0 cT = 2\pi i n \tag{15.95}$$

where n is an arbitrary integer. From (15.92), (15.94), and (15.95) we conclude

$$\alpha c \frac{T}{r} = \alpha c 2\pi \frac{\sqrt{mr}}{e} = 2\pi i n \tag{15.96}$$

Thus the possible radii r for orbits around which the length scale is preserved are given by the equation

$$r = -\frac{n^2 e^2}{\alpha^2 mc^2} \tag{15.97}$$

If we choose

$$\alpha = \frac{2\pi i e^2}{hc} = \frac{i}{137} \tag{15.98}$$

where h is Planck's constant, we obtain

$$r = n^2 \frac{h^2}{4\pi^2 e^2 m} \tag{15.99}$$

the Bohr radii of quantum theory. Thus the gauge preservation introduces automatically quantization conditions for the orbits in the hydrogen atom and an imaginary fine-structure constant.

We have been intentionally very elementary and crude in our reasoning. London showed how the general quantum conditions could be obtained in analogous fashion and pointed out the close relation between gauge theory and wave mechanics. The most unexpected feature of this argument is the fact that α is an imaginary number. If we wish to define length as a real number, this interpretation becomes difficult. On the other hand, the state vectors of quantum mechanics are complex-valued entities for which the multiplication by complex numbers is well defined and significant. Hence, after the development of modern quantum theory, Weyl interpreted the ideas of gauge invariance and the corresponding mathematical formalism as connected with transplanting the state vector of a quantum-theoretical system. Be this as it may, there seems to be a very suggestive and potentially significant content to this mathematical model. Physical reasoning led Weyl to a model of differential geometry which is of great theoretical interest and aesthetic appeal. This model, as we just showed, was soon applied in another field of physics and proved of real importance there. Therefore we found it useful to give a sketch of this attempt at unified field theory, though it is quite certain that in its present form it is unrealistic and a failure.

15.4 Some Mathematical Machinery

We have seen in the first section that general relativity theory and classical electromagnetism in the form of system (15.2) appear to be consistent and admit a reasonable solution, the Nordström metric (15.21), describing the charged point mass. In this and the following section we shall study the general system (15.2) in further detail.

Suppose we are given a source-free electromagnetic field described by the electromagnetic tensor $F_{\mu\nu}$ which gives rise to an energy-momentum tensor $T_{\mu\nu}$ in the usual way. One can then ask what is the most general tensor $R_{\mu\nu}$ which is consistent with both the equations of general relativity and Maxwell's equations (15.2). The complete answer to this question would allow us to gain insight into the relation of electromagnetism and gravitation without going beyond the present theories of Maxwell and Einstein. We would indeed have a characterization of all geometries which are possible in the presence of a pure electromagnetic field.

The first geometer to study this problem was Rainich, in 1925, soon after the advent of Einstein's general relativity theory. He obtained one condition on the metric field, described by $R_{\mu\nu}$, created by a source-free electromagnetic field (Rainich, 1925). This work was continued in 1957 by Misner and Wheeler, who gave further differential conditions on such a field. The investigations of these authors showed that a consistent "already unified" field theory could be obtained within the existing structure of electromagnetism and general relativity theory (Misner and Wheeler, 1957).

We shall approach this problem with a section of purely mathematical investigations which are also of interest in themselves, independently of the success of the Rainich-Wheeler-Misner theory. This first section will provide machinery which will enable us in the next section to derive the conditions of Rainich-Wheeler-Misner. Having arrived at that point, we shall possess the concepts needed to give a necessarily brief sketch of Wheeler's notions on the construction of the "already unified" theory of general relativity and electromagnetism.

We begin by obtaining the famous Cayley-Hamilton theorem of elementary matrix theory. Let $F = ((f_{ik}))$ be an arbitrary $n \times n$ matrix and define a series

(15.100) $$S \equiv I + tF + t^2F^2 + \cdots$$

where t is a real parameter. It is easy to see that, for sufficiently small t, this series converges absolutely, and therefore defines a new $n \times n$ matrix S. One can then verify by direct multiplication and rearrangement that

(15.101) $$(I - tF)S = S(I - tF) = I$$

Hence the inverse of $(I - tF)$, for sufficiently small t, is precisely the geometric series S:

(15.102) $$(I - tF)^{-1} = S = I + tF + t^2F^2 + \cdots$$

On the other hand, elementary matrix theory gives us an alternative way to calculate the inverse of $(I - tF)$. One knows, according to Cramer's rule, that

(15.103) $$(I - tF)^{-1}_{ik} = \frac{\beta^T{}_{ik}}{|I - tF|}$$

where β^T_{ik} is the transpose of the cofactor matrix which appears in the expansion of the determinant by minors in the ith row:

(15.104) $$|I - tF| = \sum_k (I - tF)_{ik}\beta_{ik} \qquad \text{(any } i\text{)}$$

Since β_{ik} is $(-1)^{i+k}$ times the determinant of an $(n-1) \times (n-1)$ minor of $(I - tF)$, it is clearly a polynomial of degree $n - 1$ in t. Thus (15.103) and (15.102) tell us that

(15.105) $$|I - tF|(I - tF)^{-1} = |I - tF|S = ((\beta_{ik}(t)))^T$$

is a polynomial of degree $n - 1$ in t.

We define next the characteristic polynomial of the matrix F as

(15.106) $$\phi(\lambda) = |\lambda I - F| = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$$

which is clearly a polynomial of degree n in λ. The role of this polynomial in matrix theory is well known; in particular, its roots are the eigenvalues of F. From the definition (15.106), we see immediately that

(15.107) $$|I - tF| = t^n\left|\frac{1}{t}I - F\right| = t^n\phi\left(\frac{1}{t}\right) = 1 + a_{n-1}t + \cdots + a_0t^n$$

To obtain a very interesting result, we substitute the above representation of $|I - tF|$ and the series representation (15.100) of S into Eq. (15.105). We find that the coefficient of t^n on the left side is the matrix

(15.108) $$M = F^n + a_{n-1}F^{n-1} + \cdots + a_0I$$

This may be considered as a polynomial matrix function of a matrix in the same way that an ordinary polynomial is a scalar function of a scalar. With this point of view we may indeed write $M = \phi(F)$, as is evident from (15.106) and (15.108). Since the right side of (15.105) is only of degree $n - 1$ in t, this matrix polynomial must be identically zero:

(15.109) $$\phi(F) = F^n + a_{n-1}F^{n-1} + \cdots + a_0I = 0$$

That is, the matrix F satisfies its own characteristic equation $\phi(F) = 0$. This is the fundamental Cayley-Hamilton theorem of matrix theory. It tells us in particular that every $n \times n$ matrix satisfies an algebraic equation of order n.

Since the Minkowski tensor $F_{\mu\nu}$ is antisymmetric, the special case of a 4×4 antisymmetric matrix,

$$F^T = -F \tag{15.110}$$

will be of particular interest in the next section; so we shall proceed now to investigate the characteristic polynomial, the eigenvalues, and the general form of 4×4 antisymmetric matrices. Since a transposed matrix has the same determinant as the original, we find from (15.110) that

$$\begin{aligned}\phi(-\lambda) &= |-\lambda I - F| = (-1)^n|\lambda I + F| = (-1)^n|(\lambda I + F)^T| \\ &= (-1)^n|\lambda I - F| = (-1)^n\phi(\lambda)\end{aligned} \tag{15.111}$$

Thus, for an even n, such as $n = 4$ in the present case, the characteristic polynomial is an even function of λ. From this fact and the Cayley-Hamilton theorem, it follows that F satisfies a characteristic equation in which only even powers appear:

$$F^4 + a_2F^2 + a_0I = 0 \tag{15.112}$$

Furthermore, from (15.111) we see that, if λ is a root of ϕ, $|\lambda I - F| = 0$, then $-\lambda$ is also a root, $|-\lambda I - F| = 0$. Thus the eigenvalues of a 4×4 antisymmetric matrix occur in two pairs, $\pm\lambda_1$ and $\pm\lambda_2$.

It is an easy matter, moreover, to determine the coefficients a_0 and a_2 of the characteristic polynomial in terms of the eigenvalues $\pm\lambda_1$ and $\pm\lambda_2$. Since these eigenvalues are roots of $\phi(\lambda)$, we can write $\phi(\lambda)$ as

$$\begin{aligned}\phi(\lambda) &= (\lambda - \lambda_1)(\lambda + \lambda_1)(\lambda - \lambda_2)(\lambda + \lambda_2) \\ &= \lambda^4 - (\lambda_1^2 + \lambda_2^2)\lambda^2 + \lambda_1^2\lambda_2^2\end{aligned} \tag{15.113}$$

from which it is clear that the coefficients of (15.112) are

$$\begin{aligned}a_0 &= \lambda_1^2\lambda_2^2 \\ a_2 &= -(\lambda_1^2 + \lambda_2^2)\end{aligned} \tag{15.114}$$

Therefore

$$F^4 - (\lambda_1^2 + \lambda_2^2)F^2 + \lambda_1^2\lambda_2^2 I = 0 \tag{15.115}$$

Let us now construct a 4×4 symmetric matrix

$$\Gamma = F^2 + \frac{a_2}{2} I = F^2 - \tfrac{1}{2}(\lambda_1^2 + \lambda_2^2)I \tag{15.116}$$

In the remainder of this section we shall investigate this matrix and its relation to the antisymmetric matrix F. We shall see in Sec. 15.5 that if the matrix F corresponds to the electromagnetic tensor $F_{\mu\nu}$ in a special coordinate system, then the matrix Γ is proportional to the energy-momentum tensor $T_{\mu\nu}$ of the electromagnetic field in that system. The first interesting property of Γ follows very quickly from the definition (15.116), for if we square Γ and use (15.114) and (15.112), we find that

$$\begin{aligned}\Gamma^2 &= F^4 + a_2F^2 + \frac{a_2^2}{4} I \\ &= \left(\frac{a_2^2}{4} - a_0\right) I = \tfrac{1}{4}(\lambda_1^2 - \lambda_2^2)^2 I\end{aligned} \tag{15.117}$$

That is, Γ^2 *is a multiple of the identity matrix* I. We shall assume, moreover, that $\lambda_1^2 \neq \lambda_2^2$ throughout this chapter, so that Γ^2 is not identically zero.

Another interesting property of Γ follows if we consider the Jordan canonical form of Γ. According to Jordan, any complex matrix can be transformed by a similarity transformation into a matrix of the form

$$\mathfrak{M} = \begin{pmatrix} C_1 & & & & \\ & C_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & C_N \end{pmatrix} \tag{15.118}$$

$$C_i = \begin{pmatrix} \tau_i & 1 & 0 & 0 & \cdots & 0 \\ 0 & \tau_i & 1 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & 0 & \cdots & \tau_i \end{pmatrix}$$

where C_i has τ_i's along the main diagonal, 1's along the first superdiagonal, and zeros everywhere else. That is, there exists a nonsingular Q such that

$$\Gamma = Q^{-1}\begin{pmatrix} C_1 & & & & \\ & C_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & C_N \end{pmatrix} Q \tag{15.119}$$

However, the fact that, in the present case, Γ^2 is a multiple of the identity, namely $\frac{1}{4}(\lambda_1^2 - \lambda_2^2)^2 I$, tells us that

$$\Gamma^2 = Q^{-1}\begin{pmatrix} C_1^2 & & & & \\ & C_2^2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & C_N^2 \end{pmatrix} Q = \tfrac{1}{4}(\lambda_1^2 - \lambda_2^2)^2 I \tag{15.120}$$

and hence that each C_i has the square

$$C_i^2 = \begin{pmatrix} \tau_i^2 & 2\tau_i & 1 & 0 & 0 & \cdots & 0 \\ 0 & \tau_i^2 & 2\tau_i & 1 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdots & \tau_i^2 \end{pmatrix} = \tfrac{1}{4}(\lambda_1^2 - \lambda_2^2)^2 I_i \tag{15.121}$$

This, however, is possible only if C_i is a 1×1 matrix and

$$\tau_i = \pm\tfrac{1}{2}(\lambda_1^2 - \lambda_2^2)$$

in which case the Jordan canonical form is a diagonal matrix and

$$\Gamma = Q^{-1}\begin{pmatrix} \tau_1 & & & & \\ & \tau_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \tau_n \end{pmatrix} Q \tag{15.122}$$

The τ_i are clearly the eigenvalues of Γ; we can readily obtain them from the definition of Γ in (15.116), for

$$\begin{aligned} |\tau_i I - \Gamma| &= |\tau_i I - F^2 + \tfrac{1}{2}(\lambda_1^2 + \lambda_2^2)I| \\ &= \left|\left(\tau_i + \frac{\lambda_1^2 + \lambda_2^2}{2}\right) I - F^2\right| \end{aligned} \tag{15.123}$$

Thus $\tau_i + (\lambda_1^2 + \lambda_2^2)/2$ is an eigenvalue of F^2, which we know must be either λ_1^2 or λ_2^2. Thus we see immediately that

$$\tau_1 = \tau_2 = -\tau_3 = -\tau_4 = \tau = \tfrac{1}{2}(\lambda_1^2 - \lambda_2^2) \tag{15.124}$$

and

$$\Gamma = Q^{-1}\begin{pmatrix} \tau I & 0 \\ 0 & -\tau I \end{pmatrix} Q = \tfrac{1}{2}(\lambda_1^2 - \lambda_2^2) Q^{-1}\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} Q \tag{15.125}$$

where I is the 2×2 identity matrix. From the last statement above we obtain the second interesting property of Γ; for since the trace is invariant under a similarity transformation, (15.125) tells us that Γ *has a zero trace*

$$\mathrm{Tr}\,(\Gamma) = \mathrm{Tr}\,(F^2 - \tfrac{1}{2}(\lambda_1^2 - \lambda_2^2)I) = 0 \tag{15.126}$$

From this it follows also that

$$\mathrm{Tr}\,(F^2) = 2(\lambda_1^2 - \lambda_2^2) \tag{15.127}$$

so that Γ may be written as

$$\Gamma = F^2 - \tfrac{1}{4}\mathrm{Tr}\,(F^2) I \tag{15.128}$$

in which form its null trace is manifest.

In the work of the next section we shall need the fact that the Q in (15.125) may be chosen to be an orthogonal matrix $Q^T = Q^{-1}$. Let us put this statement in the form of a theorem: *If a symmetric matrix* Γ *with eigenvalues* $\tau, \tau, -\tau,$ *and* $-\tau$ *is similar to a diagonal matrix* $\mathfrak{N}$,

$$\Gamma = Q^{-1}\begin{pmatrix} \tau I & 0 \\ 0 & -\tau I \end{pmatrix} Q = Q^{-1}\mathfrak{N} Q \tag{15.129}$$

then there exists some such Q *which is orthogonal,* $Q^{-1} = Q^T$. (The matrix Q will not necessarily be real.)

The proof of the statement is quite straightforward. Since Γ is symmetric,

$$\Gamma^T = Q^T \mathfrak{N}(Q^{-1})^T = \Gamma = Q^{-1}\mathfrak{N}Q \tag{15.130}$$

Thus

$$\mathfrak{N}(QQ^T) = (QQ^T)\mathfrak{N} \tag{15.131}$$

If we write QQ^T in terms of 2×2 blocks as

$$QQ^T = \begin{pmatrix} \alpha & \gamma \\ \delta & \beta \end{pmatrix} \tag{15.132}$$

then Eq. (15.54) tells us that

$$\begin{pmatrix} \alpha & \gamma \\ -\delta & -\beta \end{pmatrix} = \begin{pmatrix} \alpha & -\gamma \\ \delta & -\beta \end{pmatrix} \qquad \gamma = \delta = 0 \tag{15.133}$$

so we find

$$QQ^T = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \tag{15.134}$$

The submatrices α and β are clearly symmetric since QQ^T is symmetric, and moreover, since Q is nonsingular, we have the following relation between determinants:

$$|QQ^T| = |Q|^2 = |\alpha|\,|\beta| \neq 0 \tag{15.135}$$

Thus the determinants $|\alpha|$ and $|\beta|$ are both nonzero. Next note that, if Q is replaced by

$$\tilde{Q} = \begin{pmatrix} R & 0 \\ 0 & S \end{pmatrix} Q \tag{15.136}$$

with arbitrary nonsingular matrices R and S, one finds by straightforward multiplication that

$$\Gamma = Q^{-1}\mathfrak{N}Q = \tilde{Q}^{-1}\mathfrak{N}\tilde{Q} \tag{15.137}$$

and furthermore

$$\tilde{Q}\tilde{Q}^T = \begin{pmatrix} R\alpha R^T & 0 \\ 0 & S\beta S^T \end{pmatrix} \tag{15.138}$$

To complete the proof of the theorem we need only show that the equations $R\alpha R^T = I$ and $S\beta S^T = I$ have solutions R and S; then, by (15.137) and (15.138), $\tilde{Q}$ will indeed be an orthogonal matrix which puts Γ into diagonal canonical form. To solve $R\alpha R^T = I$ for R, let

$$\alpha = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \qquad R = \begin{pmatrix} u & v \\ w & z \end{pmatrix} \tag{15.139}$$

Substitution of these in $R\alpha R^T = I$ gives the following three equations in four unknowns:

$$\begin{aligned} au^2 + 2buv + cv^2 &= 1 \\ aw^2 + 2bwz + cz^2 &= 1 \\ auw + buz + bwv + czv &= 0 \end{aligned} \tag{15.140}$$

Several cases are possible for the coefficients a, b, c of these equations. Suppose, first, that $a \neq 0$, in which case the following values form a solution:

$$v = 0 \qquad u = \frac{1}{\sqrt{a}} \qquad z = \sqrt{\frac{a}{|\alpha|}} \qquad w = \frac{-b}{\sqrt{a|\alpha|}} \tag{15.141}$$

where $|\alpha|$ is the nonzero determinant of α: $ac - b^2$. The second case, $c \neq 0$, is quite similar and need not be written out. The last case, $a = c = 0$, possesses the solution

$$z = 1 \qquad u = \frac{i}{2b} \qquad v = -i \qquad w = \frac{1}{2b} \tag{15.142}$$

Thus we have shown that an R exists for which $R^T\alpha R = I$. In similar manner a solution S to $S^T\beta S = I$ also exists, so the proof of the theorem is complete.

The canonical form for Γ,

$$\Gamma = Q^T \begin{pmatrix} \tau I & 0 \\ 0 & -\tau I \end{pmatrix} Q = Q^{-1} \begin{pmatrix} \tau I & 0 \\ 0 & -\tau I \end{pmatrix} Q \tag{15.143}$$

where $\tau = \frac{1}{2}(\lambda_1^2 - \lambda_2^2)$, and Q is orthogonal, can be used to shed light on the structure of F, the matrix from which Γ is constructed. F may be written in the completely general form

$$F = Q^T \begin{pmatrix} K & L \\ M & N \end{pmatrix} Q \tag{15.144}$$

where Q is the same orthogonal matrix which appears in (15.143), and K, L, M, and N are appropriate 2×2 matrices. Note that, since F is antisymmetric, K and N are also antisymmetric. From the definition of Γ (15.116) we find that, by virtue of (15.124),

$$F^2 = \Gamma + \tfrac{1}{2}(\lambda_1^2 + \lambda_2^2)I = Q^T\begin{pmatrix}\lambda_1^2 I & 0 \\ 0 & \lambda_2^2 I\end{pmatrix}Q \tag{15.145}$$

Using this form for F^2 and (15.144) for F and the obvious identity

$$F^2F = FF^2$$

we obtain

$$\begin{pmatrix}\lambda_1^2 K & \lambda_1^2 L \\ \lambda_2^2 M & \lambda_2^2 N\end{pmatrix} = \begin{pmatrix}\lambda_1^2 K & \lambda_2^2 L \\ \lambda_1^2 M & \lambda_2^2 N\end{pmatrix} \tag{15.146}$$

By the assumption $\lambda_1^2 \neq \lambda_2^2$ (which we always make), the matrices L and M must be zero. Similarly, the identity $FF = F^2$, with the representations (15.144) and (15.145), gives

$$\begin{pmatrix}K^2 & 0 \\ 0 & N^2\end{pmatrix} = \begin{pmatrix}\lambda_1^2 I & 0 \\ 0 & \lambda_2^2 I\end{pmatrix} \tag{15.147}$$

Since K and N are 2×2 antisymmetric matrices, they must both be multiples of

$$J = \begin{pmatrix}0 & 1 \\ -1 & 0\end{pmatrix} \tag{15.148}$$

J^2 is clearly $-I$, so we see from (13.147) that

$$K = \pm i\lambda_1 J \qquad N = \pm i\lambda_2 J \tag{15.149}$$

Since the eigenvalues of F occur in pairs, $\pm\lambda_1$ and $\pm\lambda_2$, the sign in (15.149) is arbitrary, and we can simply choose the $+$. By substituting K and N from (15.149) into (15.144), we finally arrive at a general canonical form for the antisymmetric matrix F:

$$F = Q^T\begin{pmatrix}i\lambda_1 J & 0 \\ 0 & i\lambda_2 J\end{pmatrix}Q \tag{15.150}$$

with the restriction that $\lambda_1^2 \neq \lambda_2^2$.

In the rest of this section we shall not obtain any new results, but shall merely rewrite the above results in a more elegant and convenient form for use in Sec. 15.5. Define the matrices

$$p = iQ^T\begin{pmatrix}J & 0 \\ 0 & 0\end{pmatrix}Q \qquad q = iQ^T\begin{pmatrix}0 & 0 \\ 0 & J\end{pmatrix}Q \tag{15.151}$$

where Q is the same orthogonal matrix as in (15.150). Since $J^2 = -I$, we find by elementary computation that

$$\begin{gathered} p^2 + q^2 = I \qquad p^3 = p \qquad q^3 = q \\ qp = pq = 0 \qquad p^4 + q^4 = I \end{gathered} \tag{15.152}$$

The canonical form (15.150) of F now reads, in terms of p and q,

$$F = \lambda_1 p + \lambda_2 q \tag{15.153}$$

The matrix Γ can then be written with the aid of (15.152) as

$$\begin{aligned} \Gamma &= F^2 - \tfrac{1}{2}(\lambda_1^2 + \lambda_2^2)I \\ &= \lambda_1^2 p^2 + \lambda_2^2 q^2 - \tfrac{1}{2}(\lambda_1^2 + \lambda_2^2)(p^2 + q^2) \\ &= \tfrac{1}{2}(\lambda_1^2 - \lambda_2^2)(p^2 - q^2) \end{aligned} \tag{15.154}$$

which agrees with (15.125). Squaring the identity (15.154), we then obtain

$$\begin{aligned} \Gamma^2 &= \tfrac{1}{4}(\lambda_1^2 - \lambda_2^2)^2(p^4 + q^4) \\ &= \tfrac{1}{4}(\lambda_1^2 - \lambda_2^2)^2 I \\ &= \tfrac{1}{4}\mathrm{Tr}\,(\Gamma^2) I \end{aligned} \tag{15.155}$$

as in (15.128).

By comparing (15.153) and (15.155), we can obtain one last interesting result concerning the relation of F and Γ. Instead of using F to construct Γ, let us use an antisymmetric $\tilde{F}$ defined by

$$\begin{gathered} \tilde{F} = \tilde{\lambda}_1 p + \tilde{\lambda}_2 q \\ \tilde{\lambda}_1 = \cosh\alpha\sqrt{\lambda_1^2 - \lambda_2^2} \qquad \tilde{\lambda}_2 = \sinh\alpha\sqrt{\lambda_1^2 - \lambda_2^2} \end{gathered} \tag{15.156}$$

where α is an arbitrary parameter. Then, evidently,

$$\tilde{\lambda}_1^2 - \tilde{\lambda}_2^2 = \lambda_1^2 - \lambda_2^2 \tag{15.157}$$

so, by (15.155), the same Γ matrix results from both F and $\tilde{F}$:

$$\tilde{\Gamma} = \tfrac{1}{2}(\tilde{\lambda}_1^2 - \tilde{\lambda}_2^2)(p^2 - q^2) = \tfrac{1}{2}(\lambda_1^2 - \lambda_2^2)(p^2 - q^2) = \Gamma \tag{15.158}$$

Note in particular that the choice in (15.156), $\cosh \alpha = \lambda_1/\sqrt{\lambda_1^2 - \lambda_2^2}$ and $\sinh \alpha = \lambda_2/\sqrt{\lambda_1^2 - \lambda_2^2}$, clearly yields $F = \tilde{F}$. It is therefore clear that an entire one-parameter (α) family (15.156) of F matrices gives rise to the same Γ matrix.

It should be observed that the eigenvalues λ_1 and λ_2 need not be real; indeed we shall find in general that one is real and the other imaginary. If λ_1 is real and λ_2 is imaginary, we see that α must be imaginary in the case considered above. In general it is some complex-parameter field.

15.5 The Equations of Rainich, Misner, and Wheeler

We shall now proceed to apply the results of Sec. 15.4 to the task of obtaining the equations of Rainich, Misner, and Wheeler from the system (15.2). These equations involve only the contracted Riemann tensor $R_{\mu\nu}$, which describes the geometry of space-time, and constitute the basis of the already unified field theory.

For convenience we shall work in a locally geodesic system so that at some chosen fixed point P the Christoffel symbols all vanish. This makes ordinary and covariant derivatives of first order the same at P. Such a geodesic coordinate system is determined only up to a linear transformation with constant coefficients. In order to use the matrix results of Sec. 15.4, we may then use a geodesic system in which the metric tensor $g_{\mu\nu}$ at P is the Kronecker $\delta_{\mu\nu}$. That is, we shall work in a locally geodesic system with a "unit" metric tensor. One should note that even these conditions determine the system only up to an orthogonal transformation. In the system we have chosen, where the metric tensor $\delta_{\mu\nu}$ coincides with the identity matrix, the x^i coordinates will in general be imaginary and $F_{\mu\nu}$ will be complex. Furthermore, we can clearly drop the distinction between covariant and contravariant indices and consider tensors as matrices with all indices down; tensor algebra and matrix algebra at P are therefore the same.

In our system the energy-momentum tensor of the electromagnetic field

$$T_{\mu\nu} = \frac{1}{c^2}(F_\mu{}^\alpha F_{\alpha\nu} + \tfrac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}) = \frac{1}{c^2}(F_\mu{}^\alpha F_{\alpha\nu} - \tfrac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\beta\alpha}) \tag{15.159}$$

can be expressed by the matrix equation

$$c^2T = \Gamma \equiv F^2 - \tfrac{1}{4}\operatorname{Tr}(F^2)I \tag{15.160}$$

where we have introduced the new matrix $\Gamma = c^2T$ for convenience. This new matrix Γ depends on F in exactly the same way that the matrix Γ, which we studied in Sec. 15.4, depended upon the antisymmetric matrix F. Thus all the algebraic results of Sec. 15.4 can be applied to the present problem. Two facts in particular are of interest. First, Γ is traceless, which is indeed evident from (15.160):

$$\operatorname{Tr}\Gamma = 0 \tag{15.161}$$

Utilizing the proportionality between Γ and the matrix $R = R_{\mu\nu}$ as expressed by the Einstein equations in matrix form,

$$R = \frac{C}{c^2}\Gamma \tag{15.162}$$

we can rewrite (15.161) in terms of R as

$$\operatorname{Tr} R = R^\mu{}_\mu = 0 \tag{15.163}$$

Since (15.163) is written in tensor notation, it is true at all points in all coordinate systems. Second, Eq. (15.155) tells that Γ^2 is a multiple of the identity matrix I:

$$\Gamma^2 = \tfrac{1}{4}\operatorname{Tr}(\Gamma^2)I \tag{15.164}$$

Again utilizing the Einstein equations (15.162), we can easily cast this in general covariant form:

$$R_{\mu\nu}R^\nu{}_\alpha = \tfrac{1}{4}(R_{\tau\beta}R^{\tau\beta})g_{\mu\alpha} \tag{15.165}$$

Equations (15.163) and (15.165) are the first two of four sets of relations on $R_{\mu\nu}$ which form the basis of the Rainich-Wheeler-Misner theory.

It is possible to strengthen Eq. (15.165) somewhat by demonstrating that, under a reasonable physical assumption (as will be explained below), the scalar $R_{\tau\beta}R^{\tau\beta}$ is a positive real number. This is most easily shown by utilizing the expression for $\Gamma_{\tau\beta}\Gamma^{\tau\beta}$ in terms of the eigenvalues of $F_{\alpha\beta}$, that is, $(\lambda_1^2 - \lambda_2^2)^2$, as given in (15.155). An eigenvalue equation is a covariant concept. Indeed, if one expresses the eigenvalue equation of $F_{\alpha\beta}$ in the covariant form

$$F_{\alpha\beta}\xi^\beta = \lambda\xi_\alpha = \lambda g_{\alpha\beta}\xi^\beta \tag{15.166}$$

it is clear that the eigenvalue λ is indeed a scalar. Equation (15.166) gives rise in the usual way to a covariant secular equation for λ:

$$|F_{\alpha\beta} - \lambda g_{\alpha\beta}| = 0 \tag{15.167}$$

What we now wish to show is that, if we make the physically reasonable demand that $F_{\alpha\beta}$ *be real in a system of real coordinates, then* $(\lambda_1^2 - \lambda_2^2)^2$ *is a positive real number.*

A scalar can be calculated in any coordinate system, so we shall momentarily utilize a real Lorentz system with

$$g_{\alpha\beta} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \tag{15.168}$$

In this system we write $F_{\alpha\beta}$, which is now assumed to be real, in the general form

$$F_{\alpha\beta} = \begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix} \tag{15.169}$$

and $\mathbf{D} = (f,-e,d)$. The secular equation (15.167) for λ with (15.168) for $g_{\alpha\beta}$ and (15.169) for $F_{\alpha\beta}$ is found by direct expansion to be

$$\lambda^4 - (\mathbf{C}^2 - \mathbf{D}^2)\lambda^2 - (\mathbf{C} \cdot \mathbf{D})^2 = 0 \tag{15.170}$$

where we have used three-dimensional vector notation. The solutions of this quadratic equation are immediately seen to be

$$\lambda^2 = \frac{(\mathbf{C}^2 - \mathbf{D}^2) \pm \sqrt{(\mathbf{C}^2 - \mathbf{D}^2)^2 + 4(\mathbf{C} \cdot \mathbf{D})^2}}{2} \tag{15.171}$$

From this it is clear that, by (15.155),

$$(\lambda_1^2 - \lambda_2^2)^2 = (\mathbf{C}^2 - \mathbf{D}^2)^2 + 4(\mathbf{C} \cdot \mathbf{D})^2 = \Gamma_{\tau\beta}\Gamma^{\tau\beta} \tag{15.172}$$

The radical in (15.171) is clearly greater than $\mathbf{C}^2 - \mathbf{D}^2$, so that one solution for λ^2, say λ_1^2, is positive-definite and the other, say λ_2^2, is negative-definite. The two can be equal only in the so-called null case, where $\mathbf{C}^2 - \mathbf{D}^2 = \mathbf{C} \cdot \mathbf{D} = 0$. It is our standing assumption that the eigenvalues are not equal so that $(\lambda_1^2 - \lambda_2^2)^2$ is clearly a positive real number. Since it is also a scalar, we have shown that $\Gamma_{\tau\beta}\Gamma^{\tau\beta}$, and hence $R_{\tau\beta}R^{\tau\beta}$, is real and positive in general.

The above fact is interesting as a statement about the tensor $R_{\mu\nu}$, but it is also important in the derivation of a further algebraic condition on $R_{\mu\nu}$, as we shall see presently.

Recall from (10.71) that in special relativity theory the component T_{00} is proportional to the energy density of the electromagnetic field:

$$T_{00} = \frac{\mathbf{E}^2 + \mathbf{H}^2}{2c^2} \tag{15.173}$$

One would like to carry over this interpretation of T_{00} to general relativity. This clearly requires that T_{00} be of positive value in any system of real coordinates. What we must show is that $T_{00} \geq 0$, or the equivalent statement $R_{00} \leq 0$, is a covariant and consistent requirement.

To do this, consider an arbitrary vector v_α. Multiplication of v_α by $R^\alpha{}_\beta$ leads to a new vector w_β:

$$w_\beta = -R^\alpha{}_\beta v_\alpha \tag{15.174}$$

From the second algebraic condition (15.165) it is evident that the norms of w_β and v_α are proportional:

$$w_\beta w^\beta = \tfrac{1}{4}(R_{\mu\nu}R^{\mu\nu})v_\alpha v^\alpha \tag{15.175}$$

Since we have shown that $R_{\mu\nu}R^{\mu\nu}$ is a positive real number, it is clear that w_α and v_α are both timelike, both spacelike, or both null. In particular, one can therefore say that, under the linear operation (15.174), $R_{\mu\nu}$ carries the light cone into itself. This is indeed a very elegant and physically simple way of understanding the significance of the condition (15.165).

Now consider the transformation of R_{00} to a new coordinate system,

$$\bar{R}_{00} = \frac{\partial x^\mu}{\partial \bar{x}^0}\frac{\partial x^\nu}{\partial \bar{x}^0} R_{\mu\nu} \tag{15.176}$$

It is easy to see that the transformation coefficient $\partial x^\mu/\partial \bar{x}^0$ is indeed a contravariant vector, for in another primed system,

$$\frac{\partial x'^\mu}{\partial \bar{x}^0} = \frac{\partial x'^\mu}{\partial x^\lambda}\left(\frac{\partial x^\lambda}{\partial \bar{x}^0}\right) \tag{15.177}$$