INTRODUCTION TO GENERAL RELATIVITY

INTERNATIONAL SERIES IN PURE AND APPLIED PHYSICS

Adler, Bazin, and Schiffer Introduction to General Relativity
Azároff (Editor) X-Ray Spectroscopy
Azároff, Kaplow, Kato, Weiss, Wilson, and Young X-Ray Diffraction
Becker Introduction to Theoretical Mechanics
Bjorken and Drell Relativistic Quantum Fields
Bjorken and Drell Relativistic Quantum Mechanics
Chodorow and Susskind Fundamentals of Microwave Electronics
Collin Field Theory of Guided Waves
Evans The Atomic Nucleus
Fetter and Walecka Quantum Theory of Many-Particle Systems
Feynman and Hibbs Quantum Mechanics and Path Integrals
Hall Introduction to Electron Microscopy
Hardy and Perrin The Principles of Optics
Harnwell Electricity and Electromagnetism
Harnwell and Livingood Experimental Atomic Physics
Harrison Solid State Theory
Jammer The Conceptual Development of Quantum Mechanics
Krall and Trivelpiece Principles of Plasma Physics
Leighton Principles of Modern Physics
Lindsay Mechanical Radiation
Livingston and Blewett Particle Accelerators
Morse Vibration and Sound
Morse and Feshbach Methods of Theoretical Physics
Morse and Ingard Theoretical Acoustics
Newton Scattering Theory of Waves and Particles
Park Introduction to the Quantum Theory
Present Kinetic Theory of Gases
Read Dislocations in Crystals
Richtmyer, Kennard, and Cooper Introduction to Modern Physics
Rossi and Olbert Introduction to the Physics of Space
Schiff Quantum Mechanics
Schwartz Introduction to Special Relativity
Schwartz Principles of Electrodynamics
Slater Quantum Theory of Atomic Structure, Vol. I
Slater Quantum Theory of Matter
Slater Electronic Structure of Molecules: Quantum Theory of Molecules and Solids, Vol. 1
Slater Insulators, Semiconductors, and Metals: Quantum Theory of Molecules and Solids, Vol. 3
Slater The Self-consistent Field for Molecules and Solids: Quantum Theory of Molecules and Solids, Vol. 4
Slater and Frank Introduction to Theoretical Physics
Smythe Static and Dynamic Electricity
Stratton Electromagnetic Theory
Tinkham Group Theory and Quantum Mechanics
Tinkham Introduction to Superconductivity
Townes and Schawlow Microwave Spectroscopy
Wang Solid-State Electronics
White Introduction to Atomic Spectra

The late F. K. Richtmyer was Consulting Editor of the series from its inception in 1929 to his death in 1939. Lee A. DuBridge was Consulting Editor from 1939 to 1946; and G. P. Harnwell from 1947 to 1954. Leonard I. Schiff served as consultant from 1954 until his death in 1971.

INTRODUCTION TO GENERAL RELATIVITY

RONALD ADLER
Instito de Fisica
Universidade Federal de Pernambuco
Recife, Brazil

MAURICE BAZIN
Department of Physics
Rutgers University

MENAHEM SCHIFFER
Department of Mathematics
Stanford University

SECOND EDITION

McGRAW-HILL BOOK COMPANY

New York St. Louis San Francisco Auckland Düsseldorf Johannesburg Kuala Lumpur London Mexico Montreal New Delhi Panama Paris São Paulo Singapore Sydney Tokyo Toronto

INTRODUCTION TO
GENERAL RELATIVITY

1 2 3 4 5 6 7 8 9 0 KPKP 7 9 8 7 6 5

This book was set in Modern 8A by Bi-Comp, Incorporated. The editors were A. Anthony Arthur and Madelaine Eichberg; the production supervisor was Sam Ratkewitch.
Kingsport Press, Inc., was printer and binder.

Library of Congress Cataloging in Publication Data

Adler, Ronald.
Introduction to general relativity.

(International series in pure and applied physics)
1. General relativity (Physics) I. Bazin, Maurice, joint author. II. Schiffer, Menahem, joint author. III. Title.
QC173.6.A34 1975 530.1'1 74-18459
ISBN 0-07-000423-4

Contents

Preface to the Second Edition

Since the publication of the first edition of this book new developments in the theory of general relativity and in its observational and experimental verification have occurred which are so fundamental that they demand discussion even in an introductory text. Among these developments we note in particular:

1. The observational discovery of pulsars and the theoretical progress in understanding these very dense neutron stars, and the related work in understanding the endpoint of gravitational collapse known as the black hole.
2. The new or improved observational tests of general relativity such as the measured time delay of radar signals passing near the sun, and the accurately measured angular deflection of radio signals from quasars in the sun's field.
3. The scalar-tensor variant of relativity proposed by Brans and Dicke, which is viewed by some as a viable alternative to conventional relativity and by others as a foil for experimental tests.
4. Observations of new phenomena important to cosmology, such as the background electromagnetic radiation, which is interpreted as the residue of the explosive birth of the universe, referred to as the big bang.

In writing this edition we have tried to include an adequate discussion of the fundamentals of these developments while maintaining the introductory nature and clarity of the original text; we retain the philosophy of presenting a self-contained and overtly didactic text which can be read by a competent student without instruction or outside references. In particular, we hope the addition of exercises will further this aim.

The specific major changes we have made in response to the developments discussed above are as follows. Chapter 6 includes a discussion of the time delay of radar signals in a Schwarzschild field and the Kruskal coordinates which describe well the Schwarzschild geometry including the Schwarzschild sphere and its interior. Chapter 7 presents a simple derivation of the Kerr metric and a discussion of some of its properties such as the so-called spinning black-hole surface and the behavior of objects in close-in trajectories. Chapter 14 discusses stellar structure as related to relativity; very idealized systems such as a static incompressible fluid sphere and a pressureless gas sphere are studied in preference to more realistic stellar models which require much more technical astrophysics. In Chapter 11 the scalar-tensor theory of gravity is constructed as an application of variational techniques, and is discussed. Chapters 12 and 13 contain newer material on observations in cosmology, such as the background electromagnetic radiation, and an expanded discussion of the geometric interpretation of the Robertson Walker metric. Other additions are the algebraic classification of the Riemann tensor and a clarified discussion of the polarization of gravitational waves, their action on test bodies, and the problem of detection.

The exercises which have been added at the end of each chapter are intended to give students practice with the mathematical machinery and an opportunity to test their understanding of the physics. They are intended to be workable with the aid of the text alone.

The problems are to be sharply distinguished from exercises. They are intended to send students to outside references, or to provide food for thought; in general many or all will require considerable understanding and effort. We hope that these problems will provide an opportunity for students to make the transition from a classroom atmosphere to a research atmosphere by providing an acquaintance with recent literature.

We would like to thank the many people who have contributed to this second edition through thoughtful criticisms and suggestions. In particular one of us (R. J. A.) would like to acknowledge helpful discussions and correspondence with L. Caroff, D. Black, P. C. Peters, J. A. Wheeler, R. H. Dicke, J. M. Cohen, and C. Sheffield.

Ronald Adler
Maurice Bazin
Menahem Schiffer

Preface to the First Edition

The content of this book is based upon a lecture course given at Stanford University by the senior author and elaborated by his co-authors. Since there are numerous works available which deal with the general theory of relativity, some of them masterful and even classical, it seems necessary to explain the specific intention of the present book. In writing this text, our principal aim has been to show the close interaction of mathematical and physical ideas and to give the reader a feeling for the necessity and beauty of the laws of general relativity. We hope that our work will attract mathematicians to a fruitful and promising field of research which provides motivation and applications for many ideas and methods of modern analysis and differential geometry. At the same time we hope to provide the physicist with a simple and attractive introduction into powerful mathematical methods which may help him in various fields of theoretical research.

Since our main purpose in writing this book is frankly didactic, we have made a great effort to be clear and easily understood. We have tried to explain and motivate each "ansatz" even at the risk of being overly verbose and have carried out most calculations and transformations in great detail. We have preferred a lucid discussion of interesting special cases to a general and abstract formulation and have refrained from introducing mathematical concepts which may be very important in n-dimensional spaces with complex topology, but which do not have immediate applications to the physical theory considered.

Our restriction to the more elementary mathematical methods has also been motivated by the following consideration. The more elaborate the mathematical tools are, the more the future development of a physical theory is predetermined and prejudged. It seems that the theory of gravitation and of general relativity is still far from completion and may progress along lines yet unforeseen. It is therefore necessary that the basic mathematical ideas and concepts be discussed *in statu nascendi*, so that adjustments and alterations can be more easily understood and accepted if they should become necessary.

On the other hand, we wished to show how mathematical reasoning and formal simplicity lead often to selection of physical laws. To illustrate the great influence of mathematical argument in forming scientific theories, we have discussed some physical models of whose actual significance we are not quite sure, but which form an excellent proving ground for novices in these pioneering fields of theoretical physics.

We have felt that there is a great difference between the theory of special and general relativity, in concepts and in methods. In modern introductory physics courses and in lectures on electromagnetic theory the ideas and formulas of special relativity theory are well covered. Special relativity theory appears as the invariance theory of the Maxwell equations under the Lorentz group of transformations. It is now a classical and undisputed part of theoretical physics. On the other hand, general relativity theory is far from complete, is not assimilated into the mainstream of modern physical theory, and presupposes methods of differential geometry and partial differential equations which are quite specific. Hence we have decided to presuppose in this book the knowledge of the theory of special relativity theory to the extent which is usual for a present-day undergraduate student in physics. We concentrate on the actual theory of general relativity and its characteristic methods. We hope thus to have actually simplified and facilitated the approach to this theory.

We extend our thanks to Professor L. Schiff of Stanford University for encouragement toward the publication of this book. We also wish to thank Dr. R. W. Fuller, Columbia University, and Mr. G. Patsakos, Stanford University, for helpful remarks and editorial help, and Mrs. Charlotte Austin, who has given outstanding technical assistance through all stages of the manuscript.

Ronald Adler
Maurice Bazin
Menahem Schiffer

INTRODUCTION TO GENERAL RELATIVITY

Introduction

1 Physics and Geometry

The theory of general relativity which we shall describe in this book represents a fusion of mechanics and the theory of gravitation, on the one hand, and of geometry, on the other. The combination obtained will result in great formal beauty and mathematical elegance. It may therefore be hoped that the rest of physical theory will ultimately be included in the theoretical edifice which we shall develop. We shall also give some indications of how electromagnetic theory can be incorporated into the framework of general relativity theory.

The history of physics records many attempts to explain physical phenomena by geometric arguments, and the problem of space has entered the foundations of Newtonian mechanics from the beginning. We remember that the law of inertia states that a material point which is not affected by any force must move on a straight line with constant speed; that is, it must perform a uniform motion. This law is basic for discovering forces in nature. Every time that a nonuniform motion occurs in nature we are sure that forces are involved. The fact that the planets move around the sun in ellipses, for example, indicates that gravitational forces are acting. In order to apply the criteria thus provided, we have to define precisely the nature of a straight line, which is by no means an easy thing to do and leads to problems in the foundations of geometry. But worse than that, it is evident that a uniform motion relative to one observer will not be uniform for a second observer who is himself in nonuniform motion with respect to the first. Which one of the two observers has the right to claim that the law of inertia is valid in his frame of reference?

We may say in retrospect that the heliocentric theory of Copernicus leads to a reference system in which Newtonian mechanics is valid. This fact, more than the simplifications obtained for planetary theory, is the significant point in the Copernican system. It allowed the development of analytical mechanics as a beautiful and successful branch of theoretical physics and applied mathematics. Most of the scientists who developed celestial mechanics concentrated on its mathematical aspects and neglected the more philosophical question of the proper reference system. But already in Newton's time, his contemporary Leibnitz raised the question of the relativity of space and of the implications of the postulates of mechanics. For this celebrated philosopher space was nothing more than the set of possible positions of simultaneously existing bodies. Then, according to Leibnitz, space as a set of position markers could not have a physical meaning of its own, and a proper theory of mechanics should be independent of the motion of the observer relative to the purely fictitious coordinate system used to identify material points. Leibnitz pointed out that Newtonian mechanics endowed space with physical significance and that it introduced a distinguished coordinate system in which it is valid. Thus absolute space was postulated by the new theory of mechanics, and Leibnitz protested against such a concept on philosophical grounds.

Newton accepted the concept of absolute space and showed quite forcefully that physics looks quite different indeed from different reference systems. His famous pail experiment is a classic in its simplicity and convincing power. He filled a pail with water and suspended it from a twisted rope. In unwinding itself the rope set the pail in rotary motion, and the rotation of the pail continued for a while until it came to rest. The water in the pail was at rest in the first stage of the rotation of the pail and had a level surface. The fact that the pail was moving relative to it did not affect it. In the second phase of the rotation of the pail, the friction between the fluid and the wall forced the fluid to participate in the motion. Water and pail then moved as one body, and, according to Newton, the surface of the water had the form of a paraboloid of revolution due to centrifugal force on the water. In the third stage, the pail had already come to rest, but the water was still rotating. In a certain sense the situation was similar to the first stage: water and pail were in the same relative motion. But now the surface of the water was parabolic. This showed that not the relative motion of water and pail were decisive for the phenomenon of depression of the water surface, but the rotation of the body of water relative to absolute space and the consequent centrifugal force.

The Leibnitz objection was thus overruled by experiment. Space was not a mathematical scaffolding for identification of possible material occupants without any physical significance. Space exerted forces and shaped material continua. As mechanics developed, the forces created by space became better understood. Centrifugal force and Coriolis force were given simple mathematical expressions and were found in various phenomena of nature. These forces were sometimes called apparent forces, for one could select a system of reference in which they disappeared. The forces occurred only because a wrong coordinate system was used; they were the penalty for the use of an incorrect geometry.

Many philosophers of science could not get accustomed to this point of view. The apparent forces were, after all, quite real. An exploding flywheel did considerable damage even if the centrifugal force which destroyed it was only apparent. How could an abstract mental construct like geometry lead to such realistic effects? Mach (Mach, 1883) returned to the doubts of Leibnitz regarding absolute space. In his opinion it was the matter of the universe which determined geometry and the concept of an admissible frame of reference. The pail of Newton had indeed rotated against the abstract coordinate axis stuck through the center of the earth, but it was also rotated relative to all matter which fills space. If this interpretation was right, rotation relative to tremendous masses should create actual forces just like centrifugal forces. He encouraged an experiment, which was performed in 1896 by the brothers Friedländer, to measure forces on a point inside of a heavy flywheel in fast rotary motion. The experiment was negative, but this could always be attributed to the insufficiency of the masses used.

There is one criterion which distinguishes apparent forces from actual forces. Since apparent forces are all of inertial nature and since inertia is mass-proportional, apparent forces should always be mass-proportional. If one were to observe a universal effect on all bodies considered which was precisely proportional to their mass, one should then suspect that the coordinate system was wrong and that, by a proper choice of coordinates, this universal effect could be transformed away. This is surely the case for inertial force, centrifugal force, and Coriolis force.

There is another well-known universal force which affects every material point mass proportionally, namely, the force of gravity. One is not accustomed to call gravity an apparent force. But it is not difficult to show that it can indeed be transformed away by proper choice of a reference system. If we define an apparent force as a force which can be made to disappear in an appropriate coordinate system, gravitation must surely be considered as an apparent force. The transformation of gravitational forces was described by Einstein in a thought experiment which is now as classical as the Newton pail experiment. It may be considered, indeed, as a refutation of the Newton experiment in so far as it changes entirely the concept of absolute space which was derived by Newton from his experiment.

Einstein's well-known box experiment involves an observer in a closed box who feels that he and all apparatus in the box possess a mass-proportional downward acceleration. He cannot look out of the box, and he wishes to ascertain the reason for this acceleration by measurements inside. There are at least two possible interpretations: (1) There may be a heavy mass affixed to the bottom of the box producing a very nearly uniform gravitational field, and the attraction by that mass on all matter in the box may be the reason for the downward acceleration. (2) The box may be in accelerated upward motion due to a pull on a rope which is attached to the roof of the box. The downward acceleration of all matter is then nothing but the common inertia of all matter. A short consideration of this alternative will show that there is no known effect in mechanics which would allow one to distinguish between these two alternatives. The force of gravitation acting in alternative 1 may be simulated by the apparent force which accounts for the same effects in alternative 2. Hence gravitation can be transformed away, at least locally, like an inertial force and must be considered as an apparent force in the sense of our definition.

An objection might be made to Einstein's reasoning. Why should we restrict ourselves to purely mechanical measurements? Let us send a ray of light horizontally from one wall of the box to the opposite. If the box is accelerated upward, it is easy to calculate that the ray should describe a parabola and hit the opposite wall at a lower point than the point of emission. In the case of gravitation, such an effect was not foreseen by classical physics. Thus either the distinction between apparent forces and gravitation is possible by optical means, or gravitation must affect rays of light in such a way that the end result would be the same as in the case of accelerated motion. Einstein was so convinced of the validity of his thought experiment and the equivalence of the gravitational and inertial explanation that he made the daring prediction that light would be deflected in a gravitational field. It is well known that Eddington verified this hypothesis in 1919.

We see that the belief in the possibility of alternative interpretations of gravitation as an actual or inertial force leads to predictions on the behavior of nonmechanical phenomena. The axiom of indistinguishability between gravity and inertia is called the principle of equivalence. It allows many qualitative and quantitative predictions. As a matter of fact, most experimental verifications of general relativity theory which have been given up to now can be derived from skillful applications of the equivalence principle alone, without need for the more intricate and detailed formalism of the whole theory. It is also clear that many electrodynamic consequences are implicit in the equivalence principle. For example, we know much about the radiative behavior of charged accelerated conductors; the equivalence principle predicts completely analogous behavior in a gravitational field.

The equivalence principle contains a remarkable fact of physics which had not been stressed too much until the time of Einstein because it was observed at the beginning of the new mechanics and incorporated in its foundations. This fact is the equivalence of inertial mass and gravitational mass, which was established experimentally by Eötvös and more recently by Dicke. Inertial mass is a factor of proportionality which determines the acceleration of the body in question under a given external force, whatever its nature. Gravitational mass is the measure of attraction which the body exerts on a fixed test body because of the law of universal gravitation. That these two numbers which occur in so different phenomena should be identical is a surprising and unexplained fact. The equivalence principle itself does not "explain" this identity, but gives it a new and important significance.

The fundamental question now arises: Is the equivalence of the inertial and gravitational interpretation of the box experiment a formal accident, or is gravitation indeed an apparent force like centrifugal and Coriolis forces? Since it is against the spirit of the scientific method to believe easily in accidents, the choice lying at the basis of the theory of general relativity is to treat gravitation on the same footing as the classical inertial forces, the so-called apparent forces. Since these latter forces were best understood by geometric considerations, it was natural to suspect that gravitation had a closer connection with geometry than had been realized before.

Let us then analyze the axioms of classical mechanics from the point of view of a geometric interpretation. A material point which is unaffected by exterior forces moves along a straight line with constant speed This statement is valid relative to a distinguished set of reference frames which seem to have been fixed by the distribution of matter in the entire universe. So much we can accept of Mach's interpretation of inertial motion. Thus all fixed stars and galaxies of the universe determine a Euclidean geometry such that a free material point, i.e., one that is far from mass concentrations, moves along a shortest line, i.e., a geodesic or straight line. Geometry becomes a physical reality. It determines a guidance field for free mass points. The role of geodesics in analytical mechanics is well known and important. If we attach a material point to a two-dimensional surface and let it move freely on this surface without any other forces but the constraints which keep it to the surface, it will move on a geodesic of the surface. Thus, in this case, when the experimenter has prescribed the "guidance field," i.e., the surface for the particle, the geodesic motion appears again.

After these analogies it is clear how we might geometrize the theory of gravitation. A heavy attracting body, say, the sun, modifies the geometry around it in such a way that the geodesics in this geometry are the curved trajectories of the attracted particles. If we succeed in finding the law by which the matter affects geometry, the actual calculation of motion will then be reduced to the well-studied mathematical problem of determining the geodesics of a given geometry.

The calculation of geodesics is a central problem of the calculus of variations. On the other hand, variational principles have played an important role in analytical mechanics. Why has Einstein's idea of geometrizing the gravitational field of force not been conceived before? To answer this question let us look at the most geometrical of all variational principles of mechanics, namely, the principle of Maupertuis. In its simplest form it states the following: Let a particle move in a field of force with the potential $V(x,y,z)$. If it travels from a point P_1 to a point P_2 with the varying velocity v, its trajectory is that actual curve which yields a stationary value for the action integral $\int_{P_1}^{P_2} v\, ds$ among all possible paths connecting P_1 and P_2 which can be run through with the same constant energy $E = \frac{1}{2}mv^2 + V$ of the particle. We may express this principle in the obvious variational formula

$$\delta \int_{P_1}^{P_2} \left(\frac{2}{m}(E - V)\right)^{1/2} ds = 0$$

In the case of $V = 0$, we obtain the rectilinear motion asserted by the law of inertia. In the case of a nonvanishing potential $V(x,y,z)$, we can introduce a metric based on the line element

$$dl^2 = \frac{2}{m}[E - V(x,y,z)](dx_1^2 + dx_2^2 + dx_3^2)$$

and formulate the trajectory condition as

$$\delta \int_{P_1}^{P_2} dl = 0$$

In the new differential geometry with this line element dl, the trajectory would indeed be a geodesic. But observe that, for different particles in the same field and with different energies E, the geometry would have to be a different one, which is impossible. This fact precluded a geometrization of dynamics.

We can see the same difficulty from the following consideration. Suppose that the gravitational field of the sun creates a non-Euclidean geometry and that the planets have to move along the geodesics of this geometry. It is well known that, if we prescribe a point in space and a direction through this point, there exists exactly one geodesic passing through the point with the prescribed direction. On the other hand, two particles in a gravitational field fired from the same point in the same direction will move along the same trajectory only if their initial velocities are equal. Thus only one projectile could at most follow the corresponding geodesic. Indeed, geometry deals with the space variables and directions, but velocity is a concept involving time, and it is the initial velocity which enters into the determination of a trajectory. In the theory of special relativity Einstein had shown that space and time variables are inextricably connected and transform among each other under Lorentz transformations. A reduction of gravitational theory to geodesic motion in an appropriate geometry could be carried out only in the four-dimensional space-time continuum of relativity theory. That this is indeed possible is the main thesis of this book. That a reduction of the theory of gravitation to geometry was hardly possible before the special theory of relativity should be clear from the preceding considerations.

In order to achieve a geometrization of gravitation we shall have to look very closely at differential geometry and the theory of geodesics. We shall have to describe at first the conventional differential geometry as developed by Gauss and Riemann for the case of ordinary space. The mathematical formalism developed will lead to a natural extension of all concepts to the case of differential geometry in a space-time continuum as needed for the physical theory.

2 The Choice of Riemannian Geometry

The modern theory of Riemannian geometry developed from the elementary differential geometry of surfaces in Euclidean space by the usual mathematical process of abstraction. A surface in ordinary Euclidean space can be described by means of a Cartesian coordinate system in which points are characterized by their coordinates $P \equiv (x_1,x_2,x_3)$ as the locus of all points P whose coordinates satisfy an analytic relation

(1) $$f(x_1,x_2,x_3) = 0$$

Under simple assumptions on the function f, we can express the relation (1) also in the explicit form

(2) $$x_3 = \varphi(x_1,x_2)$$

which displays clearly the two-dimensional character of the surface. Indeed, the coordinates x_1, x_2 may be varied freely and the x_3 coordinate of the point P on the surface which has the projection into the plane $x_3 = 0$ with coordinates x_1, x_2 can be calculated from (2).

In order to avoid the asymmetry in the three coordinates which is displayed in (2) and to remove the restriction on f which would allow the representation (2), Gauss based his theory of surfaces on a more general parametric representation than that in (2). We introduce two parameters u_1, u_2, which can vary freely in a domain Δ of the (u_1,u_2)

plane. We choose three functions $\varphi_i(u_1,u_2)$, which are defined in Δ and are there as often differentiable as the theory we have in mind will require. By the definition

$$x_i = \varphi_i(u_1,u_2) \qquad i = 1, 2, 3 \tag{3}$$

we create a two-dimensional subset of points in the three-dimensional Euclidean space with the Cartesian coordinates x_i. This is the surface representation in Gaussian parameters. We may conceive of the u_i as coordinates of the points of the surface. We can choose for the same surface such systems of Gaussian surface coordinates in a great many ways. Indeed, if we introduce the new set of coordinates v_i by the invertible relations

$$u_i = U_i(v_1,v_2) \qquad v_i = V_i(u_1,u_2) \tag{4}$$

we may use the v_i as surface coordinates just as well as the u_i. The problem of the differential geometer is to express those laws which have an intrinsic geometric meaning in a form which is independent of the accidental choice of the surface coordinates.

As an example of such an approach, let us consider a curve on the surface described in parameter form $u_i(\tau)$. The length of such a curve between the points belonging to the parameters $\tau = 0$ and $\tau = 1$ is given by

$$\begin{aligned} L = \int_0^1 ds &= \int_0^1 \left[\left(\frac{dx_1}{d\tau}\right)^2 + \left(\frac{dx_2}{d\tau}\right)^2 + \left(\frac{dx_3}{d\tau}\right)^2\right]^{1/2} d\tau \\ &= \int_0^1 \left(\sum_{i,k=1}^{2} g_{ik}\frac{du_i}{d\tau}\frac{du_k}{d\tau}\right)^{1/2} d\tau \end{aligned} \tag{5}$$

where the g_{ik} are found by simple calculation to be

$$g_{ik} = \sum_{\nu=1}^{3} \frac{\partial x_\nu}{\partial u_i}\frac{\partial x_\nu}{\partial u_k} = \sum_{\nu=1}^{3} \frac{\partial \varphi_\nu}{\partial u_i}\frac{\partial \varphi_\nu}{\partial u_k} \tag{6}$$

In differential form we may say that the infinitesimal distance of two surface points with coordinate differences du_i is given by

$$ds = \left(\sum_{i,k=1}^{2} g_{ik}\, du_i\, du_k\right)^{1/2} \tag{7}$$

Here the g_{ik} matrix is a function of the surface coordinates u_i. If we change from the surface coordinates u_i to a new set, say, $\bar{u}_i$, the relation (7) will transform into

$$ds = \left(\sum_{i,k=1}^{2} \bar{g}_{ik}\, d\bar{u}_i\, d\bar{u}_k\right)^{1/2} \tag{7'}$$

The $\bar{g}_{ik}$ can be easily computed in terms of the g_{ik} and the transformation formulas. The formal structure of (7′) and (7) is obviously the same. Gauss showed that, from the knowledge of the so-called metric terms $g_{ik}(u_1,u_2)$, many important geometric properties of the surface could be derived.

The situation just described served as the starting point for Riemann's ideas, which he expounded in 1854 in his habilitation lecture, "On the Hypotheses Which Lie at the Foundation of Geometry" (Riemann, 1892). He pointed out that the restriction of Gauss to the case of two surface coordinates u_1, u_2 was not necessary and only motivated by the fact that Gauss had considered two-dimensional surfaces in a three-dimensional space. Riemann proposed to study the geometry of spaces where points are characterized by n coordinates $u_i (i = 1, 2, \ldots, n)$ and where the infinitesimal distance between two points with coordinate differences du_i would be given by the formula

$$ds^2 = \sum_{i,k=1}^{n} g_{ik}\, du_i\, du_k \tag{8}$$

Here the $g_{ik}(u_1, \ldots, u_n)$ should be arbitrarily prescribed functions of the coordinates u_i. However, once being given and thus determining a geometry of the space, they should transform under a transformation $u_i \leftrightarrow v_i$ of coordinates in such a way as to make ds^2 independent of the choice of coordinates used. Riemann showed that the Gaussian differential geometry could be extended in unchanged form and that concepts like curvature could be carried over into such general geometries. He also pointed out that the classical Euclidean geometry was a special case of the general theory, namely, that in which

$$ds^2 = \sum_{i=1}^{3} dx_i^2 \tag{9}$$

and that the special non-Euclidean geometry, then just recently discovered by Bolyai and Lobachevsky, entered into his larger framework.

Once we have recognized the logical possibility of replacing the Pythagorean formula (9) by the much more general formula (8) and of developing a consistent differential geometry in such spaces, we are led necessarily to the question: Why is the actual space of our experience endowed with the special metric formula (9)? Riemann conjectured that the particular choice of geometry in nature depended on the reality which created or determined space; that is, the distribution of matter and the forces acting through space should determine geometry. He ended his thesis with the statement that, at this stage, we are crossing from the field of geometry into the field of physics.

The problems raised by these deep considerations of Riemann were squarely faced by Einstein in his development of the general theory of relativity and given a solution which is logically and aesthetically satisfactory. But from the above it seems justified to consider Riemann as one of the most important precursors of modern relativity.

After Riemann had shown that the metric (9) of Euclidean geometry is a very special one and may be replaced by the more general equation (8), the question arose whether even this general form could not be further generalized. Indeed, the quadratic form on the right-hand side of (8) arose only from the fact that a two-dimensional surface had been imbedded in a three-dimensional space in which a quadratic metric form (9) was assumed to be valid. This argument does not hold in general spaces. Indeed, modern differential geometry deals with so-called Finsler spaces, in which at every point x_i of the space, the length and the differential increments are related by

$$ds = F(x_i, dx_i) \tag{10}$$

Here F is an arbitrary function of the point considered and of the differentials dx_i. It is subjected only to the natural demand of being homogeneous in the first degree in the dx_i for positive factors; i.e.,

$$F(x_i, \lambda\, dx_i) = \lambda F(x_i, dx_i) \qquad \lambda > 0 \tag{11}$$

A large body of geometric theory, in particular the theory of geodesic lines, can be extended without modification to this still more general type of metric geometry.

Why is the Finsler geometry not realized in nature? In 1868, Helmholtz published a paper entitled "On the Facts Which lie at the Foundations of Geometry" (Helmholtz, 1868). As the title indicates, it was motivated by Riemann's geometric analysis of the problem of space and was intended to complete the logical treatment by an empirical physical approach. Helmholtz pointed out that we can do certain things in real space which rule out certain logically possible geometries. For example, we can take a small rigid body, hold one of its points fixed,

and rotate the rigid body freely around this fixed point. From this fact alone most Finsler geometries are ruled out.

While this argument can be given in a space of any number of dimensions, we shall restrict ourselves to the three-dimensional space of physics for the sake of simplicity. Let O be the point held fixed in the body, and let P be a different but otherwise arbitrary point, say, at a distance 1 from O. We can first turn the body in such a way that P coincides with a given point on the unit sphere around O. Having performed such a rotation, we can hold O and P fixed and perform a second rotation of the rigid body around the axis OP with an arbitrary angle. Thus we have altogether three degrees of freedom in the rotation of a rigid body around a fixed point O. Under this three-parameter group of transformations, the mutual distances between any two points of the body are not changed. In order to abstract this situation, we start with the empirical fact that a metric function (10) exists in which such three-parameter groups of distance-preserving transformations are possible, and ask for the implications with respect to the distance concept and the ensuing differential geometry.

Suppose that near the fixed point O a metric formula

$$ds = F(dx^i) \tag{12}$$

holds, where F is a continuously differentiable function of its argument vector and is positive-homogeneous of degree 1. Clearly, this provides a local Finsler metric: we have to determine the character of $F(\xi_i)$, which has to be a consequence of our geometrical assumptions.

Introduce a local coordinate system x_i around the fixed point O and take O to be the origin of the system. We have, by assumption, the three-parameter family of transformations

$$\bar{x}_i = f_i(x_k, p_j) \tag{13}$$

which depends on the three parameters p_j. Since O is a fixed point of the transformations, we must assume that all f_i vanish at O identically in the p_j. Hence we can deduce from (13) the following linear transformation law for infinitesimal vectors dx_i:

$$d\bar{x}_i = \sum_{k=1}^{3} \alpha_{ik}(p_j)\, dx_k \tag{13$'$}$$

where the $\alpha_{ik}(p_j)$ represent the derivatives $\partial f_i/\partial x_k$ evaluated at $x_k = 0$.

We postulated the preservation of distances under the transformations considered. By this postulate, the linear transformations (13′) must satisfy the identity

$$F(d\bar{x}_i) = F(dx_i) \tag{14}$$

We can use the homogeneous character of F to rid ourselves of infinitesimals. This allows us to assert that the three-parameter group of linear transformations characterized by the matrices $((\alpha_{ik}(p_j)))$ satisfies the identity

$$F\left(\sum_{k=1}^{3} \alpha_{ik}\xi_k\right) = F(\xi_i) \tag{15}$$

Assume now that the group of linear transformations has the same freedom of operation as the group of rotations has in Euclidean geometry. That is, given an arbitrary plane in the ξ-space, there exists still a one-parameter subgroup of those transformations which carries the plane into itself. We may assume without loss of generality that the plane selected has the equation $\xi_3 = 0$. Since the transformations of our group have the form

$$\bar{\xi}_i = \sum_{k=1}^{3} \alpha_{ik}(p_1,p_2,p_3)\xi_k \tag{16}$$

we have to demand

$$\alpha_{31}(p_1,p_2,p_3) = 0 \qquad \alpha_{32}(p_1,p_2,p_3) = 0 \tag{16'}$$

which will allow us to determine p_2 and p_3 as functions of the remaining free parameter $p = p_1$.

We shall also demand that the distance function $F(\xi_i)$ be positive-definite, which is a natural requirement in view of our physical distance concept and of definition (12). From this fact, the identity (15), and the above described nature of the transformation group, we can draw far-reaching conclusions about the function $F(\xi_i)$.

Having selected the plane $\xi_3 = 0$, we consider the one-parameter subgroup of transformations (16) with the conditions (16′). This subgroup transforms each vector $(\xi_1,\xi_2,0)$ into a vector $(\bar{\xi}_1,\bar{\xi}_2,0)$; it induces in the plane $\xi_3 = 0$ a one-parameter group of linear transformations

$$\bar{\xi}_i = \sum_{k=1}^{2} \alpha_{ik}(p)\xi_k \qquad i = 1, 2 \tag{17}$$

Now let

$$F(\xi_1,\xi_2,0) = \Phi(\xi_1,\xi_2) \tag{18}$$

Then (15) implies clearly

$$\Phi(\bar{\xi}_i) = \Phi(\xi_i) \tag{19}$$

for all transformations (17). We can assume without loss of generality that $p = 0$ corresponds to the identity transformation and that the matrices depend differentiably upon the parameter p. Thus we may write (17) in the form of an expansion:

$$\bar{\xi}_i(p) = \xi_i + p\sum_{k=1}^{2} c_{ik}\xi_k + 0(p^2) \tag{17'}$$

where $((c_{ik}))$ is the derivative matrix of $((\alpha_{ik}))$ at $p = 0$. We differentiate (19) with respect to p and put $p = 0$ to find the identity

$$\sum_{i=1}^{2} \frac{\partial\Phi}{\partial\xi_i}\left(\sum_{k=1}^{2} c_{ik}\xi_k\right) = 0 \tag{20}$$

which is valid for arbitrary ξ_i. This is a partial differential equation for the function Φ, which, in turn, is the restriction of our metric function F to the plane $\xi_3 = 0$.

The implications of (20) are best discussed by studying its characteristic curves $\xi_i(t)$, where t is the curve parameter. Define $\xi_i(t)$ by the differential-equation system

$$\frac{d\xi_i}{dt} = \sum_{k=1}^{2} c_{ik}\xi_k \qquad i = 1, 2 \tag{21}$$

The curve $\xi_i(t)$ can be easily calculated from this system of two ordinary differential equations of first order since the coefficients are constant. From (20) it immediately follows that

$$\frac{d}{dt}\Phi(\xi_i(t)) = 0 \tag{22}$$

i.e., the curves obtained will be the level curves

$$\Phi(\xi_i) = \text{const} \tag{22'}$$

of the sought function Φ. We can make a qualitative statement at once about the integral curves of (21). Define

$$m = \min\Phi \qquad M = \max\Phi \tag{23}$$

on the circle $\xi_1^2 + \xi_2^2 = 1$. Because of the homogeneity of Φ we can assert that

$$mr \leq \Phi \leq Mr \qquad \text{for } \xi_1^2 + \xi_2^2 = r^2 \tag{23'}$$

Thus the curve $\Phi = a$ must lie between the circles a/M and a/m, and is thereby bounded away from both zero and infinity.

It is well known and easily seen that the general solution of (21) is of the form

$$\xi_i = A_i e^{\lambda t} + B_i e^{\mu t} \tag{24}$$

where λ and μ are the roots of the secular equation

$$\det \|c_{ik} - \lambda\delta_{ik}\| = 0 \tag{24'}$$

An integral curve which is bounded away from zero and infinity is possible only if λ and μ are pure imaginary. Hence we can bring (24) into the real form

$$\xi_i = a_i \cos \lambda t + b_i \sin \lambda t \tag{25}$$

It follows easily that the integral curve must be an ellipse. Instead of the parametric form (25), we can more conveniently write the equation of an ellipse as

$$\sum_{i,k=1}^{2} g_{ik}\xi_i\xi_k = 1 \tag{26}$$

where the g_{ik} are constant coefficients in the equation of the ellipse.

Let it be observed that the condition of the bounded integral curves poses also a very strong restriction in the coefficients c_{ik}. Indeed, only in very few cases does Eq. (24′) possess two conjugate imaginary roots. This leads to an important characterization of the group of linear transformations $((\alpha_{ik}(p_j)))$. We do not need at this point to enter into a discussion of the consequences, since we are interested only in the functions F and Φ.

We have shown that the locus Φ = const is always an ellipse in the plane $\xi_3 = 0$ with center at O. Since the ray in the ξ_3 direction can be replaced by an arbitrary ray, we have proved: The surface $F(\xi_i)$ = const is intersected by any plane through O in an ellipse.

It is now easy to conclude that $F(\xi_i)$ = const must be an ellipsoid in the three-dimensional space. Indeed, let P be a point of the surface S

which has the maximum distance from O. We may again assume that P lies on the ξ_3 axis. Consider all planes which contain the x_3 axis; they will intersect S in an ellipse which obviously must have the segment $\overline{OP}$ of the ξ_3 axis as major axis. Their minor axes will lie in the plane $\xi_3 = 0$, which is perpendicular to their common major axis. As the plane rotates around the ξ_3 axis, the minor axes will fill an ellipse in the plane $\xi_3 = 0$. Obviously, then, S is the ellipsoid whose three principal axes are the axis OP and the two principal axes of the ellipse in the plane $\xi_3 = 0$.

We know now that $F(\xi_i)$ takes constant values on the ellipsoids

$$\sum_{i,k=1}^{3} g_{ik}\xi_i\xi_k = \text{const} \tag{27}$$

Since F and the quadratic form (27) have the same surfaces of constant value, and since F is positive-homogeneous of degree 1, it is clear from elementary considerations that we must have F^2 proportional to the quadratic form (27). By appropriate adjustment of the coefficients g_{ik} we thus have the result

$$ds^2 = \Sigma g_{ik}\, dx^i\, dx^k \tag{28}$$

which indicates that the geometry is indeed Riemannian.

The preceding consideration is an example of how very decisive statements about the possible geometry can be inferred from very few facts of experience. It may be debated to what extent the existence of rigid and freely rotating bodies is guaranteed by actual experience. However, if we add to the preceding argument the principle of the simplest formal description, the choice of a Riemannian geometry in the physical world becomes quite natural.

Our point of view, then, is the following: We cover the space with a coordinate net (x_1,x_2,x_3) which serves to distinguish different points and acts as a set of point markers. The physico-geometric meaning rests not in these markers, which may be chosen very arbitrarily, but in the distance function, which is defined differentially by an equation of the form (28). The coefficient matrix g_{ik} will depend on the coordinate system in such a way as to give ds^2 a meaning independent of the markers used. We shall have to establish the transformation law for the g_{ik} under change of coordinates in such a way as to make ds^2 invariant. But beyond this covariance law of the g_{ik} will lie quantities which have ultimately to be determined by the physics of the world, in particular by the distribution of matter.

An additional complication will arise from the fact that we shall not be able to keep space and time coordinates separated in establishing a physically significant ds^2. However, our experience in a three-dimensional spatial cross section of our four-dimensional space-time world demands, by the above Helmholtz argument, that ds^2 be quadratic in the space differentials if the time-coordinate differential dt equals zero. We expect, therefore, that ds^2 will be quadratic in all space-time differentials even for an arbitrary displacement in the four-dimensional world. The only new feature of the extended space-time geometry will be the fact that ds^2 need not be positive-definite, as can indeed be inferred from the case of special relativity theory.

In order to make these general and rather abstract considerations more specific, we shall have to develop an elegant notation and proper mathematical tools, which are provided by the theory of tensor analysis. The basic problem of tensor analysis is the determination of those constructs and concepts which are independent of the accidental choice of the coordinate system employed. We shall deal in the next few chapters with mathematical results and methods to prepare the fusion of geometry and physics which will be carried out in the following part of the book.

Bibliography

Helmholtz, H. von (1868): Über die Tatsachen, die der Geometrie zu Grunde liegen, *Nachr. Ges. Wiss. Göttingen.* Reprinted in "Wissenschaftliche Abhandlungen," vol. 2, Leipzig, 1883, p. 618.

Jammer, M. (1954): "Concepts of Space," Cambridge, Mass.

Lange, L. (1886): "Die geschichtliche Entwicklung des Bewegungsbegriffes," Leipzig.

Mach, E. (1960): "Science of Mechanics," 6th ed., La Salle, Ill.

Margenau, H. (1950): "The Nature of Physical Reality," New York–Toronto–London.

Neumann, C. (1870): "Über die Prinzipien der Galilei-Newtonschen Theorie," Leipzig.

Riemann, B. (1892): "Über die Hypothesen, welche der Geometrie zu Grunde liegen, 1854," Collected Works, Leipzig. Annotated edition by H. Weyl, Berlin, 1923. Reprinted by Chelsea in "Das Kontinuum," New York, 1960.

Robertson, H. P. (1949): Geometry as a branch of physics, in P. Schilpp (ed.), "Albert Einstein, Philosopher-Scientist," Evanston, Ill., p. 313.

Schlick, M. (1920): "Space and Time in Contemporary Physics," New York.

Voss, A. (1901): Die Prinzipien der rationellen Mechanik, *Enz. math. Wiss.*, ser. A, vol. 4, no. 1, Leipzig.

Weyl, H. (1923): "Mathematische Analyse des Raumproblems," Berlin. Reprinted by Chelsea in "Das Kontinuum," New York, 1960.

Weyl, H. (1943): "Philosophy of Mathematics and Natural Science," Princeton, N.J.

Weyl, H. (1950): "Space, Time, Matter," New York.

Most general books on the theory of relativity have some discussion on the interrelation of geometry and physics.

CHAPTER

1

Tensor Algebra

In this book we shall be considering a four-dimensional space in which coordinate systems are defined in such a way that we can go from one system to another through continuous one-to-one transformations:

$$\begin{aligned} \bar{x}^j &= f^j(x^0,x^1,x^2,x^3) \\ x^k &= h^k(\bar{x}^0,\bar{x}^1,\bar{x}^2,\bar{x}^3) \end{aligned} \tag{1.1}$$

where j runs from 0 to 3 and for which the first derivatives $\partial\bar{x}^j/\partial x^i$ (i and j vary from 0 to 3) are supposed to be continuous. This implies in particular that the Jacobian of the transformation is a continuous function of position in space; moreover, the reversibility property implies that it is never zero. We therefore consider the equivalence class of coordinate systems defined through the transformation relations (1.1).

The coordinate system x^i will, in general, be defined only in a part Σ of the space considered, and the system $\bar{x}^k$ will be a marker system only in a part $\bar{\Sigma}$ of the space which may be different from Σ. We assume that Σ and $\bar{\Sigma}$ have a common part Σ_c in which the transformation relations (1.1) hold. Each coordinate neighborhood Σ may be conceived of as a local map or coordinatization of the space and the transformation law (1.1) as the law of correspondence between two different maps of overlapping regions. It is assumed that the entire space can be covered and described by such a set of overlapping coordinate neighborhoods. The neighborhood set forms an atlas of the entire space, and the transformations (1.1) allow us to proceed from sheet to sheet of the atlas. A space which allows such a covering by maps and admits a Riemannian metric is called a Riemannian manifold. (For a discussion of the physical construction of a Riemannian manifold in astronomy see Sec. 12.1.)

We shall deal throughout with a Riemann space with a metric

$$ds^2 = \sum_{ik} g_{ik}\, dx^i\, dx^k \tag{1.2}$$

We assume that the matrix of coefficients g_{ik} is symmetric ($g_{ik} = g_{ki}$) and has the signature $(+1, -1, -1, -1)$. By definition such a metric will be called a hyperbolic metric. (These numbers denote the signs of the eigenvalues of the symmetric matrix g_{ik}; see also Sec. 5.6.)

Many theorems in this chapter will be proved for n-dimensional spaces, but we shall always try to keep in contact with physics through the concepts of general relativity for which one uses only a four-dimensional space.

1.1 Definition of Scalars, Contravariant Vectors, and Covariant Vectors

To comply with the postulates of the theory of general relativity which requires that physical laws be invariant under any change of coordinate system as defined previously, we shall look for mathematical entities which possess certain invariance properties under an arbitrary change of coordinates. Particular invariance properties will be illustrated with the simplest quantities which are defined so far over the space by the given coordinate systems and the given metric.

Scalar quantities. A scalar quantity, or in short a scalar, is a quantity which can be "measured with a scale." It is a number and does not depend on the choice of a frame of reference.

A scalar field is a point function in the space considered. It may be numerically assigned, as, for example, the temperature distribution in space-time, or may be expressed as an analytic expression of coordinate-dependent quantities, with the property of remaining invariant under a change of coordinates. A local scalar may depend on the local values of such a scalar field.

The square of the line element ds of the space is such an invariant quantity under any change of coordinates since we assume that it has a physical meaning as a space-time distance between two infinitely close world events. Thus $ds^2 = \sum_{ik} g_{ik}\, dx^i\, dx^k$ has to keep the same numerical value under an arbitrary change of coordinates: it is a scalar. The quantities dx^i and g_{ik} in turn will have to transform under a change of coordinates in such a way as to leave ds^2 invariant.

Contravariant vectors. Consider an infinitesimal displacement in space from a point A labeled by the markers (x^i) to a point B labeled by $(x^i + dx^i)$ in a given coordinate system. Let us see what the coordinate differentials dx^i which represent the infinitesimal displacement in the original coordinate system (x) become in another coordinate system $(\bar{x})$. By differentiation of Eqs. (1.1) we obtain immediately

$$d\bar{x}^i = \sum_{j=0}^{3} \frac{\partial f^i}{\partial x^j}\, dx^j$$

which we can write

$$d\bar{x}^i = \sum_{j=0}^{3} \frac{\partial \bar{x}^i}{\partial x^j}\, dx^j \tag{1.3}$$

We shall say that any set of four quantities ξ^i ($i = 0, 1, 2, 3$) which transform according to (1.3),

$$\bar{\xi}^i = \sum_{j=0}^{3} \frac{\partial \bar{x}^i}{\partial x^j}\, \xi^j$$

forms a contravariant vector. By convention, we denote the components of a contravariant vector by an index written above the letter representing the vector.

Let ξ^i and η^i be two arbitrary contravariant vectors at a given point x^i. Then the sum $a\xi^i + b\eta^i$ ($i = 0, 1, 2, 3$) will also be a contravariant vector if a and b are scalars. This follows from the homogeneous linear character of the transformation laws.

Covariant vectors. Consider a point M in space defined by the coordinates x^i in a particular coordinate system. Consider a function $\varphi(x^i)$ of the point M in space and defined in a neighborhood of M; being a function of a point, its value does not change whichever coordinate system one uses. It is therefore what we have called a scalar field.

Consider now how the four quantities $A_i = \partial\varphi/\partial x^i$ transform under a change of coordinate system from (x^i) to $(\bar{x}^i)$. The four corresponding quantities in the barred coordinate system become, by the rules of differentiation,

$$\frac{\partial \varphi}{\partial \bar{x}^i} = \sum_{j=0}^{3} \frac{\partial x^j}{\partial \bar{x}^i} \frac{\partial \varphi}{\partial x^j}$$

which gives

$$\bar{A}_i = \sum_{j} \frac{\partial x^j}{\partial \bar{x}^i}\, A_j \tag{1.4}$$

We now call a covariant vector any set of quantities transforming according to (1.4). We denote the components of a covariant vector by an index written below the letter representing the vector.

Theorem. The product $\sum_i A_i \xi^i$ formed from a covariant vector and a contravariant vector is a scalar invariant.

From (1.3) and (1.4) the product $\bar{P} = \sum_i \bar{A}_i \bar{\xi}^i$ is equal to

$$\sum_i \sum_j \sum_k \frac{\partial x^k}{\partial \bar{x}^i} A_k \frac{\partial \bar{x}^i}{\partial x^j} \xi^j$$

By differentiation of the formulas of coordinate transformation (1.1), written as identities $x^k = h^k[\bar{x}^i(x^j)]$, we have

$$\frac{\partial x^k}{\partial x^j} = \delta^k_j = \sum_i \frac{\partial h^k}{\partial \bar{x}^i} \frac{\partial \bar{x}^i}{\partial x^j} = \sum_i \frac{\partial x^k}{\partial \bar{x}^i} \frac{\partial \bar{x}^i}{\partial x^j}$$

where δ^k_j is the familiar Kronecker symbol: $\delta^k_j = 0$ for $j \neq k$; $\delta^k_j = 1$ for $j = k$. Thus

(1.5) $$\sum_i \frac{\partial x^k}{\partial \bar{x}^i} \frac{\partial \bar{x}^i}{\partial x^j} = \delta^k_j$$

and we obtain for $\bar{P}$

$$\bar{P} = \sum_{jk} \delta^k_j A_k \xi^j = \sum_j A_j \xi^j$$

If we write

(1.6) $$P = \sum_j A_j \xi^j$$

this expression is formally identical with the original quantity $\bar{P} = \sum_i \bar{A}_i \bar{\xi}^i$, and therefore P is a scalar. In analogy with ordinary vector algebra, we shall call P the *inner product* (or scalar product) of a covariant with a contravariant vector.

1.2 Einstein's Summation Convention

We notice that, in formulas (1.3) and (1.4), the summation is performed over the index j, which always occurs twice in the formula. One can therefore introduce a simplifying convention, which Einstein described in the following terms to his friend L. Kollros: "I made a great discovery in mathematics; I suppressed the summation sign every time that the summation has to be done on an index which appears twice in the general term" (Kollros, 1956).

Instead of $\bar{A}_i = \sum_j (\partial x^j/\partial \bar{x}^i) A_j$, we shall consistently write in the future $\bar{A}_i = (\partial x^j/\partial \bar{x}^i) A_j$. The index j over which one sums is called a *dummy index* because the way one labels it is irrelevant; obviously, one can write

$$\bar{A}_i = \frac{\partial x^j}{\partial \bar{x}^i} A_j = \frac{\partial x^k}{\partial \bar{x}^i} A_k$$

The justification of the above convention resides in the fact that the two sides of an equation must be indexed in the same way to represent quantities of the same character (covariant or contravariant). Therefore when more indices appear on one side than on the other, some of them must be dummy indices and be summed over. These will usually appear in pairs of one upper and one lower index. We shall use Einstein's convention only in this case. If, exceptionally, indices in the same position (for instance, both lower) are to be summed over in an expression, we shall keep the summation sign. (The reader should note that certain authors extend the use of Einstein's convention to such a case.)

1.3 Definitions of Tensors

Intrinsic definition. We shall here consider a vector in an n-dimensional space as a set of n indexed numbers which obey a given transformation law:

$$\bar{\xi}^i = \frac{\partial \bar{x}^i}{\partial x^j} \xi^j \qquad \text{for contravariant vectors (indices above)}$$

$$\bar{\eta}_i = \frac{\partial x^j}{\partial \bar{x}^i} \eta_j \qquad \text{for covariant vectors (indices below)}$$

Consider a multilinear form P:

(1.7) $$P = (T^{i_1 i_2 \cdots i_a}_{j_1 j_2 \cdots j_b})(\xi^{j_1}_{(1)} \xi^{j_2}_{(2)} \cdots \xi^{j_b}_{(b)})(\eta^{(1)}_{i_1} \eta^{(2)}_{i_2} \cdots \eta^{(a)}_{i_a})$$

where $\xi^{j_1}_{(1)}$ = j_1th component of an *arbitrary* contravariant vector labeled (1), etc.

$\eta^{(1)}_{i_1}$ = i_1th component of an *arbitrary* covariant vector labeled (1), etc.

and where $T^{i_1i_2\cdots i_a}_{j_1j_2\cdots j_b}$ is a set of n^{a+b} elements with **a** upper and **b** lower indices which will be called contravariant and covariant, respectively. We call a product $\xi^{j_1}_{(1)}\xi^{j_2}_{(2)} \cdots \xi^{j_b}_{(b)}\eta^{(1)}_{i_1}\eta^{(2)}_{i_2} \cdots \eta^{(a)}_{i_a}$ of components of arbitrary vectors an arbitrary multinomial.

By definition we say that the quantities

$$T^{i_1i_2\cdots i_a}_{j_1j_2\cdots j_b}$$

form the components of a tensor if, for an arbitrary change of coordinates under which the vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ transform according to the laws given above, they transform in such a way that P remains unchanged (is a scalar). $T^{i_1i_2\cdots i_a}_{j_1j_2\cdots j_b}$ will be called a tensor of rank $(a + b)$ with **a** contravariant indices and **b** covariant indices.

From now on we shall adopt the following notation: A tensor will be denoted either by its components, "the tensor $T^{\alpha\beta}_{\gamma}$," or for brevity by a single letter representing the set of all components which build up the tensor, "the tensor **T**."

Particular cases

1. A tensor of rank zero is a scalar, a number which is invariant under a change of coordinates. For instance, a function of a point $\varphi(x^i)$ is a scalar, as we have seen earlier.

2. A tensor of rank 1 can be either contravariant, T^i, or covariant, T_i. Let us consider, for instance, the case of a contravariant tensor T^i. Following our intrinsic definition of a tensor given above, we take an arbitrary covariant vector η_i and form $P = T^i\eta_i$, which must be a scalar. This we know is true if T^i is a contravariant vector from the theorem in Sec. 1.1. Therefore a contravariant vector is a tensor of rank 1 with one contravariant index. Similarly, a covariant vector is a covariant tensor of rank 1. We shall see later that the converse of this property is true by the quotient theorem (Sec. 1.7).

3. Let us consider the g_{ik} coefficients which define the metric in a Riemann space. We have

$$g_{ik}\,dx^i\,dx^k = ds^2 = \text{(invariant length)} \tag{1.8}$$

Our intrinsic definition would suggest that g_{ik} are components of a covariant tensor of rank 2. However, the intrinsic definition of a tensor as given in Sec. 1.3 by means of the multilinear form

$$P = T_{ik}\xi^i_{(1)}\xi^k_{(2)}$$

states that P must be a scalar quantity when $\boldsymbol{\xi}_{(1)}$ and $\boldsymbol{\xi}_{(2)}$ are two *arbitrary* vectors. Here dx^i and dx^k are different components of the *same* contravariant vector **dx**. But the proof that g_{ik} is a tensor remains valid in this case since **dx** is an arbitrary vector which one can take as the sum of two arbitrary vectors $\mathbf{dx}_{(1)}$ and $\mathbf{dx}_{(2)}$. One has, then,

$$\begin{aligned} ds^2 &= g_{ik}(dx^i_{(1)} + dx^i_{(2)})(dx^k_{(1)} + dx^k_{(2)}) \\ &= g_{ik}\,dx^i_{(1)}\,dx^k_{(1)} + g_{ik}\,dx^i_{(2)}\,dx^k_{(2)} + 2g_{ik}\,dx^i_{(1)}\,dx^k_{(2)} \end{aligned}$$

In this expression the two first terms are the scalars $ds^2_{(1)}$ and $ds^2_{(2)}$, and therefore the third term is a scalar too, which proves that g_{ik} is a covariant tensor according to our definition.

General transformation law of a tensor. We wish to derive from the intrinsic definition the general transformation law of tensors. We take our definition above and write out the requirement of invariance under an arbitrary change of coordinates:

$$(\bar{T}^{i_1\cdots i_a}_{j_1\cdots j_b})(\bar{\xi}^{j_1}_{(1)} \cdots \bar{\xi}^{j_b}_{(b)})(\bar{\eta}^{(1)}_{i_1} \cdots \bar{\eta}^{(a)}_{i_a}) = (T^{i_1\cdots i_a}_{j_1\cdots j_b})(\xi^{j_1}_{(1)} \cdots \xi^{j_b}_{(b)})(\eta^{(1)}_{i_1} \cdots \eta^{(a)}_{i_a})$$

In the left-hand side we express the barred vector components as a function of the unbarred ones and identify the coefficients of the arbitrary multinomials on both sides. We obtain

$$\bar{T}^{i_1\cdots i_a}_{j_1\cdots j_b}\left(\frac{\partial\bar{x}^{j_1}}{\partial x^{\beta_1}} \cdots \frac{\partial\bar{x}^{j_b}}{\partial x^{\beta_b}}\right)\left(\frac{\partial x^{\alpha_1}}{\partial\bar{x}^{i_1}} \cdots \frac{\partial x^{\alpha_a}}{\partial\bar{x}^{i_a}}\right) = T^{\alpha_1\cdots\alpha_a}_{\beta_1\cdots\beta_b} \tag{1.9}$$

We wish to invert this formula to obtain the barred components as a function of the unbarred ones. By multiplying both sides by

$$\left(\frac{\partial x^{\beta_1}}{\partial\bar{x}^{l_1}} \cdots \frac{\partial x^{\beta_b}}{\partial\bar{x}^{l_b}}\right)\left(\frac{\partial\bar{x}^{k_1}}{\partial x^{\alpha_1}} \cdots \frac{\partial\bar{x}^{k_a}}{\partial x^{\alpha_a}}\right)$$

we obtain terms in the left-hand side like $\dfrac{\partial\bar{x}^{j_1}}{\partial x^{\beta_1}}\dfrac{\partial x^{\beta_1}}{\partial\bar{x}^{l_1}}$, which we know are equal to $\delta^{j_1}_{l_1}$. Therefore we obtain, finally,

$$\bar{T}^{k_1\cdots k_a}_{l_1\cdots l_b} = \frac{\partial\bar{x}^{k_1}}{\partial x^{\alpha_1}} \cdots \frac{\partial\bar{x}^{k_a}}{\partial x^{\alpha_a}}\frac{\partial x^{\beta_1}}{\partial\bar{x}^{l_1}} \cdots \frac{\partial x^{\beta_b}}{\partial\bar{x}^{l_b}}T^{\alpha_1\cdots\alpha_a}_{\beta_1\cdots\beta_b} \tag{1.10}$$

One notices that this inversion could have been obtained from (1.9) by exchanging the roles of the barred and unbarred coordinates in our head and relabeling the dummy indices; this is in fact a consequence of the symmetry and self-consistency of our original definition.

Let us write out formula (1.10) for a particular case, since this is the form we shall use most often:

$$\bar{T}^{\alpha\beta}_{\gamma} = \frac{\partial \bar{x}^\alpha}{\partial x^i}\frac{\partial \bar{x}^\beta}{\partial x^j}\frac{\partial x^k}{\partial \bar{x}^\gamma} T^{ij}_k$$

The transformation law (1.10) is the general *axiomatic definition of a tensor*. It is obvious that this definition and the intrinsic definition from which we started are equivalent.

1.4 Tensor Algebra

Equality of tensors. Two tensors **A** and **B** will be called equal if their components are equal, i.e., if

$$A^{\alpha\beta}_{\gamma} = B^{\alpha\beta}_{\gamma}$$

for all values of the indices.

Important Remark. It is not necessary to specify that the components of the two tensors have to be equal in all coordinate systems. It is sufficient to know that both **A** and **B** are tensors and that their components are equal in one particular coordinate system; their components will then be equal in any coordinate system. This is obviously true from the axiomatic definition (1.10) of a tensor.

This remark will be of great help in the next chapters in finding the form of certain tensor expressions in an arbitrary coordinate system when they are known in a particular one. Also, to prove an identity between quantities which are known to be tensors, one can choose a particular coordinate system in which the identity has a simple form and is therefore easier to prove. (This will be illustrated by the repeated use of geodesic coordinate systems in Sec. 3.2.)

Properties. The definition (1.10) is linear and homogeneous in the tensor components $T^{\alpha_1\cdots\alpha_a}_{\beta_1\cdots\beta_b}$. Therefore

1. The sum of two tensors with the same number of covariant and contravariant indices can be defined as the sum of their components and is again a tensor. For example,

$$A^{\alpha\beta}_{\gamma} + B^{\alpha\beta}_{\gamma} = C^{\alpha\beta}_{\gamma}$$

2. The product of a tensor by a scalar (multiplication of each component by the scalar) is again a tensor.

Tensor multiplication. We shall avoid here general notations and consider a particular case. All the reasoning will be applicable in the general case of tensors of arbitrary rank. If $T^{\alpha\beta}_{\gamma}$ and $S^{\mu\nu}$ are two given tensors, consider the quantities

$$G^{\alpha\beta\mu\nu}_{\gamma} = T^{\alpha\beta}_{\gamma} S^{\mu\nu}$$

We say that $G^{\alpha\beta\mu\nu}_{\gamma}$ is a tensor four times contravariant and once covariant. To show this, consider the quantities

$$\bar{G}^{\alpha\beta\mu\nu}_{\gamma} = \bar{T}^{\alpha\beta}_{\gamma} \bar{S}^{\mu\nu}$$

in another coordinate system (denoted by a bar) and apply the known transformation law to $\bar{T}^{\alpha\beta}_{\gamma}$ and $\bar{S}^{\mu\nu}$:

$$\bar{T}^{\alpha\beta}_{\gamma}\bar{S}^{\mu\nu} = \frac{\partial \bar{x}^\alpha}{\partial x^i}\frac{\partial \bar{x}^\beta}{\partial x^j}\frac{\partial x^k}{\partial \bar{x}^\gamma} T^{ij}_k \frac{\partial \bar{x}^\mu}{\partial x^l}\frac{\partial \bar{x}^\nu}{\partial x^m} S^{lm}$$

Using the definition of $G^{\alpha\beta\mu\nu}_{\gamma}$ given above, we obtain

$$\bar{T}^{\alpha\beta}_{\gamma}\bar{S}^{\mu\nu} = \frac{\partial \bar{x}^\alpha}{\partial x^i}\frac{\partial \bar{x}^\beta}{\partial x^j}\frac{\partial x^k}{\partial \bar{x}^\gamma}\frac{\partial \bar{x}^\mu}{\partial x^l}\frac{\partial \bar{x}^\nu}{\partial x^m} G^{ijlm}_k$$

This shows that

$$\bar{G}^{\alpha\beta\mu\nu}_{\gamma} = \frac{\partial \bar{x}^\alpha}{\partial x^i}\frac{\partial \bar{x}^\beta}{\partial x^j}\frac{\partial \bar{x}^\mu}{\partial x^l}\frac{\partial \bar{x}^\nu}{\partial x^m}\frac{\partial x^k}{\partial \bar{x}^\gamma} G^{ijlm}_k$$

and G^{ijlm}_k is therefore a tensor by our axiomatic definition.

The *tensor product* $T^{\alpha\beta}_{\gamma}S^{\mu\nu} = G^{\alpha\beta\mu\nu}_{\gamma}$ is often called the *outer product* of the two tensors $T^{\alpha\beta}_{\gamma}$ and $S^{\mu\nu}$.

Example. Consider three vectors ξ^i, η^k, ζ^l and consider the quantities $t^{ikl} = \xi^i\eta^k\zeta^l$, which are the components of the tensor product of the three vectors. By the property shown above, they are components of a tensor of rank 3.

We shall show in the next section that, conversely, any tensor can be written as a sum of vector products.

1.5 Decomposition of a Tensor into a Sum of Vector Products (Tensor Products of Tensors of Rank 1)

Theorem. In an n-dimensional space, any tensor of rank $q > 1$ can be written as the sum of tensor products of vectors with q factors each. n^{q-1} is in general the minimum number of vector products into which a tensor can be decomposed.

The first statement in this theorem is of great importance in establishing other theorems in tensor algebra. The second statement, that n^{q-1} is the *minimum* number of terms, is of mathematical interest in its own right, although of lesser importance in further developments of tensor algebra. To make the statement of the theorem clear, let us take the example of a four-dimensional space and a contravariant tensor of rank 2: $T^{\mu\nu}$ (μ, $\nu = 0, 1, 2, 3$). The theorem states that one can write $T^{\mu\nu}$ as a sum of $4^{2-1} = 4$ tensor products of vectors:

$$T^{\mu\nu} = A^{\mu}_{(1)}B^{\nu}_{(1)} + A^{\mu}_{(2)}B^{\nu}_{(2)} + A^{\mu}_{(3)}B^{\nu}_{(3)} + A^{\mu}_{(4)}B^{\nu}_{(4)} \tag{1.11}$$

where the indices in parentheses label the different vectors.

If we have a tensor with mixed indices, the vectors will have to be either contravariant or covariant in the straightforward manner indicated below. Consider, for example, a mixed tensor of rank 3: $T^{\mu\nu}_{\sigma}$ with $n = 4$, $q = 3$. The theorem states that it can be written as the sum of $4^{3-1} = 16$ tensor products of vectors:

$$T^{\mu\nu}_{\sigma} = A^{\mu}_{(1)}B^{\nu}_{(1)}C_{(1)\sigma} + A^{\mu}_{(2)}B^{\nu}_{(2)}C_{(2)\sigma} + \cdots + A^{\mu}_{(16)}B^{\nu}_{(16)}C_{(16)\sigma} \tag{1.12}$$

It is obvious a priori that one can achieve such a decomposition with n^q vector products since this represents the number of independent components of a tensor of rank q in an n-dimensional space. The purpose of the present theorem is to establish a minimum number of such vector products.

Proof. It is sufficient to first prove the theorem in a particular coordinate system. For when the quantities $A^{\mu}_{(i)}$, $B^{\nu}_{(i)}$, $C_{(i)\sigma}$ are found in one coordinate system, we can then consider them as vector components, and by the law of tensor products (Sec. 1.4) and the addition property, the right-hand side of (1.12) will transform like a tensor and will therefore remain equal to the original $T^{\mu\nu}_{\sigma}$ tensor in all coordinate systems.

The proof will be made by induction on the rank q of the tensor. In order to simplify the writing of indices, we shall deal with contravariant indices only and with dimension n equal to 4. The proof remains applicable to any number of covariant and contravariant indices and any dimension.

We start by proving the theorem for $q = 2$, which is expressed by formula (1.11). Let us first consider the terms for which $\mu = 0$. For those terms, (1.11) will be true if we can solve the system

$$T^{00} = A^{0}_{(1)}B^{0}_{(1)} + A^{0}_{(2)}B^{0}_{(2)} + A^{0}_{(3)}B^{0}_{(3)} + A^{0}_{(4)}B^{0}_{(4)}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$T^{03} = A^{0}_{(1)}B^{3}_{(1)} + \cdots + \cdots + A^{0}_{(4)}B^{3}_{(4)}$$

Consider the four coefficients $A^{0}_{(1)}$, $A^{0}_{(2)}$, $A^{0}_{(3)}$, and $A^{0}_{(4)}$ as unknown; then this is a system of four linear equations with four unknowns. Therefore, if we choose the four vectors $\mathbf{B}_{(i)}$ arbitrarily [but with no four of them in a three-plane to avoid a vanishing determinant, say, $B^{\alpha}_{(i)} = \delta^{\alpha}_{i}$], we can solve for the 0th components of the vectors $\mathbf{A}_{(i)}$.

If we then consider the other values of μ ($\mu = 1, 2, 3$), keeping our choice of the $\mathbf{B}_{(i)}$ vectors the same, we can determine all components of the $\mathbf{A}_{(i)}$ vectors. This proves the theorem for $q = 2$. To complete the proof for an arbitrary rank, we need only show that if we admit the theorem as true for the rank $q - 1$, it is true for the rank q.

Let us consider a tensor of rank q written in a particular coordinate system and let us single out the qth index by writing the $q - 1$ first indices together inside a bracket: $T^{[\alpha\cdots\delta]\gamma}$. We now consider the indices in the bracket as being fixed; let us solve the system of four equations

$$\begin{aligned} T^{[\alpha\cdots\delta]0} &= S^{[\alpha\cdots\delta]}_{(1)}C^{0}_{(1)} + \cdots + S^{[\alpha\cdots\delta]}_{(4)}C^{0}_{(4)} \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ T^{[\alpha\cdots\delta]3} &= S^{[\alpha\cdots\delta]}_{(1)}C^{3}_{(1)} + \cdots + S^{[\alpha\cdots\delta]}_{(4)}C^{3}_{(4)} \end{aligned} \tag{1.13}$$

where the four coefficients $S^{[\alpha\cdots\delta]}_{(i)}$ ($i = 1, 2, 3, 4$) are taken as unknown. Choosing the quantities $C^{\gamma}_{(i)}$ arbitrarily (forming a nonvanishing determinant), the four unknowns are determined uniquely. Each of the four quantities like $S^{[\alpha\cdots\delta]}_{(i)}$ possesses $q - 1$ indices and, by assumption in our inductive reasoning, can be decomposed into a sum of $4^{(q-1)-1} = 4^{q-2}$ terms consisting of $q - 1$ vector products each. In order to solve the system (1.13) to get the rank q, we introduced four new vectors $\mathbf{C}_{(i)}$. Therefore the number of terms in the decomposition of a tensor of rank q becomes $4 \times 4^{q-2} = 4^{q-1}$, which completes the proof of the first part of the theorem. Note that our method of proof provides also a recursive construction procedure for the asserted tensor representation.

We now have to show that n^{q-1} is in general the minimum possible number of terms in the decomposition. To prove this, let us show that we cannot introduce less than four independent vectors at each step of the inductive reasoning.

Consider the system (1.13), which we have to solve, and suppose we try to introduce only three vectors $\mathbf{C}$: $\mathbf{C}_{(1)}$, $\mathbf{C}_{(2)}$, $\mathbf{C}_{(3)}$; thus we have only three quantities S: $S^{[\alpha\cdots\delta]}_{(1)}$, $S^{[\alpha\cdots\delta]}_{(2)}$, $S^{[\alpha\cdots\delta]}_{(3)}$. The system (1.13) becomes a system of four equations with three unknown for each value of the indices in the bracket. This system is compatible if the determinants (one for each possible combination of the indices $[\alpha\ \cdots\ \delta]$)

$$\begin{vmatrix} T^{[\alpha\cdots\delta]0} & C^0_{(1)} & C^0_{(2)} & C^0_{(3)} \\ \cdots & \cdot & \cdot & \cdot \\ \cdots & \cdot & \cdot & \cdot \\ \cdots & \cdot & \cdot & \cdot \\ T^{[\alpha\cdots\delta]3} & C^3_{(1)} & C^3_{(2)} & C^3_{(3)} \end{vmatrix}$$

are zero. To interpret these conditions, let us consider their geometrical meaning in four-space. The 4^{q-1} vectors $T^{[\alpha\cdots\delta]}$ of components $T^{[\alpha\cdots\delta]0}$, $T^{[\alpha\cdots\delta]1}$, $T^{[\alpha\cdots\delta]2}$, $T^{[\alpha\cdots\delta]3}$ must all be in the three-plane, determined by the three vectors $\mathbf{C}_{(1)}$, $\mathbf{C}_{(2)}$, $\mathbf{C}_{(3)}$, and therefore cannot be independent vectors, which is in contradiction to the fact that the original tensor is given with arbitrary components. This proves that the minimum number of vectors that must be introduced at each step of the inductive reasoning is four. The above proof is obviously applicable to the very first step of the induction process, namely, for $q = 2$, when there is only one index in the square bracket.

Therefore we have proved that 4^{q-1} is the minimum number of vector products into which a tensor of rank q can be decomposed in the way illustrated by (1.12).

1.6 Contraction of Indices

We have seen earlier how to create from given tensors new tensors of higher rank by tensor multiplication. In this section we shall show that one can create from a given tensor new tensors of lower rank through the contraction operation. Consider a tensor of rank $a + b$, $T^{i_1 i_2\cdots i_a}_{j_1 j_2\cdots j_b}$, and set $i_a = j_b = \sigma$. This gives $T^{i_1 i_2\cdots i_{a-1}\sigma}_{j_1 j_2\cdots j_{b-1}\sigma}$, which by Einstein's convention means the sum of such terms over all values of σ.

Theorem. $T^{i_1 i_2\cdots i_{a-1}\sigma}_{j_1 j_2\cdots j_{b-1}\sigma}$ is a tensor of rank $a + b - 2$, which one can denote $R^{i_1 i_2\cdots i_{a-1}}_{j_1 j_2\cdots j_{b-1}}$.

Proof. Let us write the transformation law of the tensor $T^{i_1 i_2\cdots i_a}_{j_1 j_2\cdots j_b}$ and then make $i_a = j_b = \sigma$. This gives

$$\bar T^{i_1 i_2\cdots i_{a-1}\sigma}_{j_1 j_2\cdots j_{b-1}\sigma} = \frac{\partial \bar x^{i_1}}{\partial x^{\alpha_1}} \cdots \frac{\partial \bar x^{i_{a-1}}}{\partial x^{\alpha_{a-1}}}\left(\frac{\partial \bar x^{\sigma}}{\partial x^{\alpha_a}}\right)\frac{\partial x^{\beta_1}}{\partial \bar x^{j_1}} \cdots \frac{\partial x^{\beta_{b-1}}}{\partial \bar x^{j_{b-1}}}\left(\frac{\partial x^{\beta_b}}{\partial \bar x^{\sigma}}\right) T^{\alpha_1\cdots\alpha_a}_{\beta_1\cdots\beta_b}$$

Grouping the two terms within parentheses, we get, by (1.5),

$$\frac{\partial \bar x^{\sigma}}{\partial x^{\alpha_a}}\frac{\partial x^{\beta_b}}{\partial \bar x^{\sigma}} = \frac{\partial x^{\beta_b}}{\partial x^{\alpha_a}} = \delta^{\beta_b}_{\alpha_a}$$

This forces us to write $\beta_b = \alpha_a = t$ in the right-hand side of the above equation. Thus

$$\bar T^{i_1 i_2\cdots i_{a-1}\sigma}_{j_1 j_2\cdots j_{b-1}\sigma} = \frac{\partial \bar x^{i_1}}{\partial x^{\alpha_1}} \cdots \frac{\partial \bar x^{i_{a-1}}}{\partial x^{\alpha_{a-1}}}\frac{\partial x^{\beta_1}}{\partial \bar x^{j_1}} \cdots \frac{\partial x^{\beta_{b-1}}}{\partial \bar x^{j_{b-1}}} T^{\alpha_1\alpha_2\cdots\alpha_{a-1}t}_{\beta_1\beta_2\cdots\beta_{b-1}t}$$

which shows that $T^{\alpha_1\alpha_2\cdots\alpha_{a-1}t}_{\beta_1\beta_2\cdots\beta_{b-1}t}$ is a tensor of rank

$$(a - 1) + (b - 1) = a + b - 2$$

which we can call $R^{i_1 i_2\cdots i_{a-1}}_{j_1 j_2\cdots j_{b-1}}$. This contraction operation, which creates new tensors out of a given tensor with mixed indices (covariant and contravariant), is called in German "rejuvenation" of the tensor (*Verjüngung*)!

1.7 The Quotient Theorem

We possess, so far, two criteria to recognize that a certain matrix is a tensor: the definition of a tensor by its transformation property (1.9) and the intrinsic definition with which we started our development (Sec. 1.3). We shall introduce here several tests of the tensor character of a given matrix; they could each be considered as an intrinsic definition of tensors.

Special case. Let us first treat a special case. Suppose $T^{i_1\cdots i_p}_{j_1\cdots j_r}$ is a matrix (given with a certain transformation law to express it in different coordinate systems) and ξ^{j_r} is an arbitrary vector. Suppose also that it is known that

$$S^{i_1\cdots i_p}_{j_1\cdots j_{r-1}} = T^{i_1\cdots i_p}_{j_1\cdots j_r}\xi^{j_r} \tag{1.14}$$

is a tensor; then the theorem states that $T^{i_1\cdots i_p}_{j_1\cdots j_r}$ is a tensor.

Proof. Multiply both sides of Eq. (1.14) by p arbitrary covariant vectors $\eta^{(1)}_{i_1}\ \cdots\ \eta^{(p)}_{i_p}$ and $(r - 1)$ arbitrary contravariant vectors $\xi^{j_1}_{(1)}\ \cdots\ \xi^{j_{r-1}}_{(r-1)}$:

$$S^{i_1\cdots i_p}_{j_1\cdots j_{r-1}}\eta^{(1)}_{i_1}\cdots\eta^{(p)}_{i_p}\xi^{j_1}_{(1)}\cdots\xi^{j_{r-1}}_{(r-1)} = T^{i_1\cdots i_p}_{j_1\cdots j_r}\xi^{j_r}\xi^{j_{r-1}}_{(r-1)}\cdots\xi^{j_1}_{(1)}\eta^{(p)}_{i_p}\cdots\eta^{(1)}_{i_1}$$

The left-hand side is a scalar from the intrinsic definition (Sec. 1.3) applied to the tensor S; therefore so is the right-hand side. But **T** is then clearly a tensor, as is evident from the intrinsic definition of a tensor, since ξ^{j_r} itself is arbitrary by hypothesis. This proves the theorem in the special case.

General case. We now turn to the general case, when the arbitrary vector ξ is replaced by a tensor. Suppose $T^{i_1 \cdots i_p}_{j_1 \cdots j_r}$ is a given matrix and $A^{j_l \cdots j_r}_{i_k \cdots i_p}$ is an arbitrary tensor. Suppose also it is known that

$$S^{i_1 \cdots i_{k-1}}_{j_1 \cdots j_{l-1}} = T^{i_1 \cdots i_{k-1} i_k \cdots i_p}_{j_1 \cdots j_{l-1} j_l \cdots j_r} A^{j_l \cdots j_r}_{i_k \cdots i_p} \tag{1.15}$$

is a tensor; then the theorem states that $T^{i_1 \cdots i_p}_{j_1 \cdots j_p}$ is a tensor.

Proof. Consider the arbitrary tensor **A** as a sum of vector products $\xi^{j_l}_{(l)} \cdots \xi^{j_r}_{(r)} \eta^{(k)}_{i_k} \cdots \eta^{(p)}_{i_p}$. We then multiply both sides of (1.15) by arbitrary vectors $\xi^{j_1}_{(1)} \cdots \xi^{j_{l-1}}_{(l-1)}$ and $\eta^{(1)}_{i_1} \cdots \eta^{(k-1)}_{i_{k-1}}$, and a reasoning analogous to the one used for the special case above applies here.

1.8 Lowering and Raising of Indices—Associated Tensors

Lowering of indices. Let us consider a tensor $T^{\alpha\beta}$, twice contravariant, and form its tensor product with a symmetric second-rank covariant tensor $g_{\gamma\beta}$, contracting the index β at the same time. We define the tensor obtained:

$$T^{\alpha}{}_{\gamma} = g_{\gamma\beta} T^{\alpha\beta}$$

This is a mixed tensor of rank 2. We call $T^{\alpha}{}_{\gamma}$ a mixed tensor associated with the tensor $T^{\alpha\beta}$. The operation of contracted multiplication with $g_{\gamma\beta}$ has, in a sense, "lowered" the index β of the original tensor. The notation $T^{\alpha}{}_{\gamma}$ preserves the order of the indices in $T^{\alpha\beta}$ (α first); without this convention one mixed component would represent two quantities which are in general unequal unless $T^{\alpha\beta} = T^{\beta\alpha}$. That is,

$$T^3_1 = g_{1\beta} T^{3\beta}$$

and

$$T^3_1 = g_{1\beta} T^{\beta 3}$$

are in general unequal. In our notation they are $T^3{}_1$ and $T_1{}^3$, respectively. We can repeat the operation and form the tensor

$$T_{\delta\gamma} = g_{\alpha\delta} T^{\alpha}{}_{\gamma}$$

and call it the twice-covariant tensor associated with $T^{\alpha\beta}$. We therefore write by definition

$$\begin{aligned} T^{\alpha}{}_{\gamma} &= g_{\gamma\beta} T^{\alpha\beta} \\ T_{\delta\gamma} &= g_{\alpha\delta} T^{\alpha}{}_{\gamma} = g_{\alpha\delta} g_{\gamma\beta} T^{\alpha\beta} \end{aligned} \tag{1.16}$$

The extension of this procedure to an arbitrary number of indices is obvious. In this way we can lower as many contravariant indices of a tensor of any rank as we wish and create associated mixed and covariant tensors. Note that the second-rank tensor g_{ik} which we are using to lower indices can be chosen arbitrarily. However, once selected, it plays a central role in tensor calculus since it establishes a relation between contravariant and covariant tensors; it is called the fundamental tensor. In a metric space, such as the four-dimensional space of general relativity, it is quite natural to take for g_{ik} the metric tensor itself, which is thus often called the fundamental tensor. This particular choice of g_{ik} will be seen to be most convenient in Chap. 3, when we shall consider the covariant differentiation of tensors.

Raising of indices. We should ask ourselves if a procedure analogous to the one used above can be found to raise indices. This can be done if we define a contravariant tensor that will play the role which the tensor g_{ik} had in lowering indices. For this purpose let us consider in *one particular coordinate system* the inverse matrix of the matrix of the g_{ik} coefficients. Its coefficients are

$$g^{ik} = \frac{\Delta^{ik}}{g} \qquad \Delta^{ik} = \text{cofactor of } g_{ik},\ g = \text{determinant of } g_{ik} \text{ matrix}$$

They are completely characterized by the property

$$g^{ik} g_{jk} = \delta^i_j \tag{1.17}$$

We now define a tensor g^{ik} by applying to the matrix g^{ik}, known in one coordinate system, the transformation law of tensors. Thus, going from a coordinate system (x) to a coordinate system $(\bar{x})$, we define

$$\bar{g}^{\alpha\beta} = \frac{\partial \bar{x}^\alpha}{\partial x^i} \frac{\partial \bar{x}^\beta}{\partial x^k} g^{ik}$$

Since g_{ik} is a tensor, our definition of g^{ik} as a tensor will be compatible with the original definition (1.17) if this definition itself is invariant under a change of coordinates. From the property of tensor products, the right-hand side of (1.17) must therefore be a tensor. Thus we need verify only that the matrix of Kronecker symbols is a tensor. This is immediately seen by considering the scalar product of two arbitrary vectors ξ^i and η_k, which can be written by definition of δ^i_j as

$$\xi^i\eta_i = (\delta^i_j\xi^j)\eta_i$$

In this equation $\xi^i\eta_i$ is a scalar; $\delta^i_j\xi^j$ is a vector by the quotient theorem since η_i is arbitrary. Since ξ^j is also an arbitrary vector, we can apply the quotient theorem again to prove that δ^i_j is a tensor.

It should be noted that, from the definition (1.17), the fundamental tensor with mixed indices is the unit tensor

$$g^j_i = \delta^j_i$$

With the help of the tensor g^{ik} we can raise indices in exactly the same way in which we lowered them with g_{ik}; with the tensor $T_{\gamma\delta}$ we associate the tensors $T_\gamma{}^\alpha$ and $T^{\beta\alpha}$ defined by

$$\begin{aligned} T_\gamma{}^\alpha &= g^{\alpha\delta}T_{\gamma\delta} \\ T^{\beta\alpha} &= g^{\gamma\beta}T_\gamma{}^\alpha = g^{\gamma\beta}g^{\alpha\delta}T_{\gamma\delta} \end{aligned} \tag{1.18}$$

Comparing (1.18) and (1.16), we see that the concept of associated tensors is reciprocal; this is also apparent in the definition (1.17).

We shall think of such tensors associated through contracted multiplication with the fundamental tensor or its inverse as being different mathematical aspects of one given geometrical or physical entity. This is why we denote associated tensors by the same letter **T**. To stress the underlying unity, we shall call the different aspects (contravariant, covariant, mixed) of a tensor its various *representations*. We know that, according to Einstein's postulate of covariance, all physical equations must be tensor equations. By repeated contracted multiplication with the fundamental tensor, one obtains the same equations between the components of the different representations of the tensor involved, but the equations represent always the same geometrical or physical relationship.

1.9 Connection with Vector Calculus in Euclidean Space

In this chapter we have defined the contravariant and covariant components of a tensor axiomatically by giving their transformation properties (1.10) under a change of coordinate system. We shall give here a geometrical illustration of contravariant and covariant representations of a vector in the familiar case of a Euclidean space; associated tensors of

rank 1 will appear as two different ways of defining components of the same geometrical vector. Historically, the notion of contravariant and covariant representations was naturally introduced by generalizing this particular case of vectors in a Euclidean space, and not in the axiomatic way which we chose to follow in this chapter.

Let us, for example, consider a vector **v** in a two-dimensional Euclidean vector space. Consider a coordinate system defined by the unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Decomposing **v** along such vectors gives (see Fig. 1.1)

$$\mathbf{v} = \lambda^1\mathbf{e}_1 + \lambda^2\mathbf{e}_2 \tag{1.19a}$$

λ^1 and λ^2 are called contravariant components of **v**. Projecting **v** orthogonally on the directions of $\mathbf{e}_1$ and $\mathbf{e}_2$, we obtain the two projections λ_1 and λ_2:

$$\begin{aligned} \lambda_1 &= \mathbf{v}\cdot\mathbf{e}_1 \\ \lambda_2 &= \mathbf{v}\cdot\mathbf{e}_2 \end{aligned} \tag{1.20a}$$

Here the dot represents the well-known scalar product in Euclidean space. The vector **v** is obviously completely defined by the values of these projections; they are called covariant components of **v**. Let us now consider the scalar product defined in this Euclidean vector space between two vectors **u** and **v**. We shall find out how the scalar product $\mathbf{u}\cdot\mathbf{v}$ can be expressed in terms of covariant and contravariant components. These will then appear as having the properties of what we called before covariant and contravariant vectors in the Riemann space considered.

We define the two kinds of components of **u** by

(1.19*b*) Contravariant (μ^1,μ^2) $\quad \mathbf{u} = \mu^1\mathbf{e}_1 + \mu^2\mathbf{e}_2$

(1.20*b*) Covariant (μ_1,μ_2) $\quad \mu_1 = \mathbf{u}\cdot\mathbf{e}_1, \quad \mu_2 = \mathbf{u}\cdot\mathbf{e}_2$

Let us write out the scalar product $\mathbf{u}\cdot\mathbf{v}$:

$$\mathbf{u}\cdot\mathbf{v} = (\mu^1\mathbf{e}_1 + \mu^2\mathbf{e}_2)\cdot(\lambda^1\mathbf{e}_1 + \lambda^2\mathbf{e}_2)$$

Using the associative and distributive properties of the scalar product, we obtain

$$\mathbf{u}\cdot\mathbf{v} = \mu^1\lambda^1\mathbf{e}_1\cdot\mathbf{e}_1 + \mu^1\lambda^2\mathbf{e}_1\cdot\mathbf{e}_2 + \mu^2\lambda^1\mathbf{e}_2\cdot\mathbf{e}_1 + \mu^2\lambda^2\mathbf{e}_2\cdot\mathbf{e}_2$$

If we designate by g_{ij} the scalar products $\mathbf{e}_i\cdot\mathbf{e}_j$ $(i, j = 1, 2)$ of the base vectors, we have

$$u \cdot v = g_{ij}\lambda^i\mu^j \tag{1.21}$$

where $g_{ij} = g_{ji}$ from the commutative property of the scalar product. But we could write $\mathbf{u} \cdot \mathbf{v}$ in another form:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot (\lambda^1\mathbf{e}_1 + \lambda^2\mathbf{e}^2) = \lambda^1\mathbf{u} \cdot \mathbf{e}_1 + \lambda^2\mathbf{u} \cdot \mathbf{e}_2$$

By the definition (1.20*b*) of the covariant components of $\mathbf{u}$, this becomes

$$\mathbf{u} \cdot \mathbf{v} = \lambda^1\mu_1 + \lambda^2\mu_2$$

or in tensor notation,

$$\mathbf{u} \cdot \mathbf{v} = \lambda^i\mu_i \tag{1.22}$$

Since the vector $\mathbf{v}$ can be chosen arbitrarily, formulas (1.21) and (1.22) imply that

$$\mu_i = g_{ij}\mu^j \tag{1.23}$$

Formulas (1.21) to (1.23) are reminiscent of the tensor formulas (1.8), (1.6), (1.16), which we had earlier. Let us prove that they are indeed tensor formulas, namely, that they keep their form after a change of coordinate system. This is immediately evident if we think of the geometrical meaning of Eqs. (1.21) and (1.22) in Euclidean space. First of all, the scalar product $\mathbf{u} \cdot \mathbf{v}$ depends only on the two vectors $\mathbf{u}$ and $\mathbf{v}$ and is therefore a scalar invariant under a change of coordinates. With new base vectors $\bar{\mathbf{e}}_1$ and $\bar{\mathbf{e}}_2$, we define geometrically the contravariant and covariant components $\bar{\lambda}^i$, $\bar{\mu}^i$ and $\bar{\lambda}_i$, $\bar{\mu}_i$ of $\mathbf{u}$ and $\mathbf{v}$. These correspond to λ^i, μ^i and λ_i, μ_i, respectively. In this new reference frame the scalar product $\mathbf{u} \cdot \mathbf{v}$ still has the form

$$\mathbf{u} \cdot \mathbf{v} = \bar{g}_{ik}\bar{\lambda}^i\bar{\mu}^k$$

which shows the tensor character of (1.21). The same geometrical interpretation can be applied to investigate (1.22), and this gives

$$\mathbf{u} \cdot \mathbf{v} = \lambda^i\mu_i = \bar{\lambda}^i\bar{\mu}_i$$

In the case of a vector in Euclidean space, formula (1.23) gives the geometrical interpretation of the concept of associated tensors which we introduced in the previous section: Associated vectors correspond to the covariant and contravariant components of a given geometrical vector.

By considering an infinitesimal vector $\mathbf{u}$ of length ds and contravariant components dx^i, we can write the metric of the Euclidean space

$$ds^2 = g_{ij}\, dx^i\, dx^j$$

Here we know the geometrical interpretation of the coefficients g_{ij}:

$$g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$$

In differential geometry in a metric space, for instance on a surface on which

$$ds^2 = g_{ij}\, dx^i\, dx^j$$

where dx^i and dx^j are coordinate increments on the surface, we can also interpret the g_{ij} coefficients locally as scalar products of the unit tangent vectors along coordinate lines. Such base vectors define locally a *tangent Euclidean space.*

1.10 Connection between Bilinear Forms and Tensor Calculus

We shall consider certain properties of bilinear forms under linear transformations on the variables. This will lead to the notion of contravariant relationship and thus indicate the origin of the nomenclature "contravariant" and "covariant" for the components of a vector. More generally, one could study properties of multilinear forms and show their connection with the contravariant and covariant character of tensor components in general.

Restricting ourselves to bilinear forms, we consider two sets of variables:

1. A set of r quantities X_i
2. A set of r quantities Y_i

One should note that these two sets are simply sets of arbitrary indexed quantities and that the index position has no relevance here. We therefore write all indices in the same place, and do not use Einstein's convention. Then

1. To the set X_i we apply a linear transformation with a matrix $\mathbf{A}$ of coefficients a_{ki}.
2. To the set Y_i we apply a linear transformation with a matrix $\mathbf{B}$ of coefficients b_{ki}.

These transformations give two new sets of quantities:

$$\bar{X}_k = \sum_i a_{ki} X_i \qquad \bar{Y}_k = \sum_i b_{ki} Y_i \tag{1.24}$$

Consider now the bilinear form

$$F = \sum_i X_i Y_i$$

$$\bar{F} = \sum_k \bar{X}_k \bar{Y}_k$$

Let us find the condition which **A** and **B** must satisfy in order that $F = \bar{F}$, that is, such that the form F remains formally invariant under the transformation of variables (1.24). To do this, we put the values given by (1.24) into

$$\bar{F} = \sum_k \bar{X}_k \bar{Y}_k$$

which becomes

$$\bar{F} = \sum_{ijk} a_{ki} X_i b_{kj} Y_j = \sum_{ij} \left(\sum_k a_{ki} b_{kj}\right) X_i Y_j$$

We shall have $\bar{F} = F$ as desired if

$$\sum_k a_{ki} b_{kj} = \delta_{ij} \tag{1.25}$$

Denoting by $\mathbf{A}^T$ the matrix of coefficients a_{ik} (the transpose of the matrix **A** of coefficients a_{ki}), we may write (1.25) in matrix notation as

$$\mathbf{A}^T\mathbf{B} = \mathbf{I} \qquad (\mathbf{I} \text{ is the unit matrix}) \tag{1.26}$$

or equivalently,

$$\mathbf{A}\mathbf{B}^T = \mathbf{I}$$

Matrices **A** and **B**, which are linked by the relation (1.26), are said to be *contragredient* to each other. Each matrix **A** with a nonvanishing determinant has a unique contragredient **B**, and this relation is reciprocal:

$$\mathbf{B} = (\mathbf{A}^T)^{-1}$$

implies

$$\mathbf{A} = (\mathbf{B}^T)^{-1}$$

One also sees that the contragredient matrix of the contragredient matrix of **A** is **A** itself (the contragredience relationship is an involution):

Original matrix	$\mathbf{A}$
Matrix contragredient to **A**	$\mathbf{B} = (\mathbf{A}^T)^{-1}$
Matrix contragredient to **B**	$\mathbf{C} = (\mathbf{B}^T)^{-1} = (\mathbf{B}^{-1})^T = (\mathbf{A}^T)^T = \mathbf{A}$

Furthermore, the contragredience relationship is an automorphism; i.e., it preserves the law of multiplication with the order of the factors:

If	$\mathbf{A} = (\mathbf{B}^T)^{-1}$
and	$\mathbf{D} = (\mathbf{E}^T)^{-1}$
then	$\mathbf{AD} = [(\mathbf{BE})^T]^{-1}$

When (1.26) holds, one says that the two transformations in (1.24) represented by the matrices **A** and **B**, respectively, are contragredient to one another and that the two sets of variables X and Y transform contragrediently to one another, or that they are in *contravariant* relationship.

Applications. Consider in a Euclidean vector space a vector

$$\mathbf{v} = \sum_i x^i \mathbf{e}_i$$

where the $\mathbf{e}_i$ are base vectors of the space, and apply a linear transformation on the vectors $\mathbf{e}_i$. The quantities x^i will have to transform contravariantly to the $\mathbf{e}_i$ so that **v** remains the same ($\mathbf{e}_i$ plays here the role of Y_i and x^i of X_i). Therefore the contravariant components of a vector, as defined earlier, transform contragrediently to the base vectors; hence their name.

Consider now the covariant components of a vector **u** in a Euclidean space.

$$\mu_i = \mathbf{u} \cdot \mathbf{e}_i$$

If we require that the scalar product

$$u \cdot v = \sum_i \mu_i x^i$$

If we require that the scalar product

$$u \cdot v = \sum_i \mu_i x^i$$

be unchanged by a linear transformation on $\mathbf{e}_i$ which induces a linear transformation on x^i, μ_i must transform contragrediently to x^i. But since x^i itself transforms contragrediently to $\mathbf{e}_i$, as we have seen above, μ_i must have the same matrix of transformation as $\mathbf{e}_i$. One says that the μ_i's transform cogrediently to the base vectors; hence the μ_i's are called covariant components.

Exercises

1.1 Show that if $S^{ij} = S^{ji}$ is a symmetric tensor and $A^{ij} = -A^{ji}$ is an antisymmetric tensor, then the scalar $A^{ij}S_{ij}$ is identically zero.

1.2 Show that an arbitrary second-rank tensor may be written as the sum of a symmetric tensor and an antisymmetric tensor. Now show that if one writes the line element ds^2 with g_{ik} expressed as such a sum, the contribution of the antisymmetric part of g_{ik} is zero.

1.3 In many areas of physics one deals with linear coordinate transformations of the form $\bar{x}^i = a^i_j x^j$, where the a^i_j are constant, e.g., the theory of rotations and special relativity. Show that for this special case the coordinates themselves transform as contravariant vectors.

1.4 Show that "contraction" over two contravariant or two covariant indices of a tensor does not give another tensor. For example $\Sigma_\alpha T_{\alpha\alpha}$ is not a tensor.

1.5 Verify that in three dimensions the cosine of the angle θ between vectors **A** and **B** may be written in manifestly invariant form as

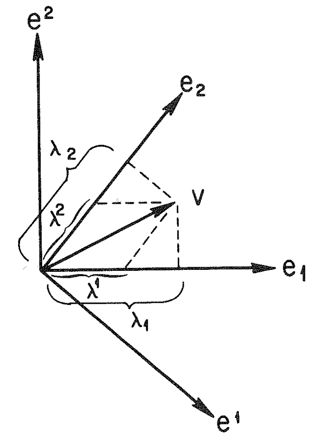
$$\mathbf{A} \cdot \mathbf{B} = \frac{A^i B_i}{\sqrt{A^l A_l B^k B_k}}$$

Show that this can be used as a generalized definition of an angle in an arbitrary n-dimensional space only if the metric g_{ij} is a positive-definite matrix. Show that this is equivalent to the signature's being $(1,1, \ldots ,1)$.

1.6 What is g_{il} for Cartesian coordinates in Euclidean three-dimensional space? On the two-dimensional surface of a sphere, $x^2 + y^2 + z^2 = R^2$, what is the metric in terms of the coordinates x and y? Repeat the exercise for cylindrical and spherical coordinates and note the natural advantage of the spherical coordinates in describing the spherical surface.

1.7 In the preceding problem we obtained the metric for cylindrical coordinates (ρ,θ,z) in ordinary three-dimensional Euclidean space. On a two-dimensional surface defined by $z = f(\rho)$, where ρ is the radial coordinate, obtain the metric in terms of ρ and θ. Choose a specific function to describe a "mountain" and compare with the "flat" space defined by $z =$ const.

1.8 The basis vectors $\mathbf{e}_i$ discussed in Sec. 1.9 have lower indices. A vector **V** is then expressible as in (1.19*a*). We could also express the vector in terms of basis vectors with upper indices, $\mathbf{e}^l$. These are defined to obey $\mathbf{e}^l \cdot \mathbf{e}_i = g^l_i = \delta^l_i$. Draw the $\mathbf{e}_i$ and $\mathbf{e}^l$ in a diagram for two dimensions. Show, in the notation of Sec. 1.9, that $\mathbf{V} = \lambda_i \mathbf{e}^i$ and $g^{ij} = \mathbf{e}^i \cdot \mathbf{e}^j$ (see Fig. 1.1).



Bibliography

Brand, L. (1947): "Vector and Tensor Analysis," New York.
Kollros, L. (1956): Albert Einstein en Suisse, Souvenirs, in "Jubilee of Relativity Theory," A. Mercier and M. Kervaire (eds.), Basel, Switzerland.
Lichnerowicz, A. (1951): "Éléments du calcul tensoriel," 2d ed., Paris.
Michal, A. D. (1937): "Matrix and Tensor Calculus," London.
Misner, C. W., K. S. Thorne, and J. A. Wheeler (1973): "Gravitation," San Francisco.
Reichardt, H. (1957): "Vorlesungen über Vektor- und Tensorrechnung," Berlin.
Rindler, W. (1969): "Essential Relativity," New York.
Synge, J. L., and A. Schild (1952): "Tensor Calculus," 2d ed., Toronto.
Weinberg, S. (1972): "Gravitation and Cosmology," New York.

See also Bibliography to Chaps. 2 and 3.

CHAPTER

2

Vector Fields in Affine and Riemann Space

In the first chapter we introduced the concept of a tensor and defined algebraic operations on tensors. But we dealt almost entirely with tensors attached to one given point of the space, and all the operations we defined involved only tensors attached to the same point. We now wish to introduce the concept of a tensor field in order to be able to define a way of comparing tensors at different points in space.

Definition. A tensor field consists of the assignment of a tensor to each point of the space; one considers the components of a tensor to be functions of the point of attachment which is characterized by coordinates or markers x^ι. We shall assume that the components of the tensor field are twice-differentiable functions of the coordinates.

Let us begin this investigation with the special case of tensors of rank 1 (vectors). In Euclidean geometry, when using rectilinear coordinate systems, we know that equality of two vectors at two *different* points can be simply identified with equality of the components of the vectors at these two points regardless of how far apart the two points actually are; for instance, in classical mechanics we are accustomed to speak of two forces as being equal although they are applied at different points. More important, a force-free motion in classical mechanics is characterized by the fact that the momentum vector remains constant; more generally, the force vector is measured by the rate of change of the momentum vector. We thus speak of a constant vector field if the vector components are constant all over the space; we shall see that it is not possible to carry this definition over to Riemann spaces because, according to such a definition, "equal" vectors would not be equally long.

In a Riemann space the notion of constancy of a tensor field remains to be defined.

2.1 Vector Transplantation and Affine Connections

Let us first consider the intuitive attempt to define a constant vector field in a Riemann space in terms of constancy of components. It will be evident that such a definition is in general not consistent with equality of length of the two vectors at the two different points.

Let us ask in what kind of Riemann space this elementary definition could hold. To answer this question we consider a particular vector field

$$\xi^{(1)}(P) = (dx^1, 0, \ldots, 0)$$

with the same constant first component dx^1 at each point of the space. At a point P the length of the vector is $g_{11}(P)(dx^1)^2$, while at another point Q it is $g_{11}(Q)(dx^1)^2$. These two lengths are not equal unless $g_{11}(P) = g_{11}(Q)$, which means that g_{11} must be constant over the space. Consider now another vector field

$$\xi^{(i)}(P) = (0, 0, \ldots, dx^i, 0, \ldots, 0)$$

with the same ith component all over the space; the reasoning above shows that g_{ii} must be a constant over the space. Then taking $\xi^{(i)} + \xi^{(j)}$ as our field, we should find that g_{ij} must be constant too. Thus the definition of a constant vector field through constancy of the components throughout the space requires us to be able to find one coordinate system in which all g_{ik}'s are constant in the large. A space which possesses this property is called a *pseudo-Euclidean space*, and the particular coordinate system in which the definition applies is rectilinear. Indeed, by a linear transformation with constant coefficients we may bring the metric form into a simple canonical form (e.g., a diagonal matrix with ± 1 components) for all points of the coordinate space. Therefore the definition of a constant vector field in terms of constancy of components cannot apply to more general types of spaces.

It is not actually necessary to introduce the comparison of lengths (a metric property) to discard the definition of a constant vector field in terms of constancy of components. It is readily seen that this definition is not coordinate-invariant, and therefore is not consistent with tensor notation. Indeed, let us take a vector field ξ^i with constant components in a given coordinate system (x^α). Changing to another arbitrary coordinate system $(\bar{x}^\alpha)$, we obtain

$$\bar{\xi}^i = \frac{\partial \bar{x}^i}{\partial x^k}\,\xi^k$$

Thus $\bar{\xi}^i$ is not a constant over the space since $\partial\bar{x}^i/\partial x^k$ is arbitrary. The only case in which such a definition is acceptable is that in which we restrict ourselves to a linear transformation of coordinates such as occurs in particular for a change of rectilinear coordinates in a pseudo-Euclidean space.

We shall now investigate further the matter of covariance requirements on the definition of a constant vector field. We wish to define the constancy of a vector field intrinsically and independently of a coordinate system; furthermore, the definition must contain the usual Euclidean definition of constancy as a particular case. Let us therefore see how we could modify the elementary definition we have just rejected. With this aim we shall investigate the behavior under coordinate transformations of a vector field which has constant components in one particular coordinate system. Starting with a vector field ξ^i of constant components in an original coordinate system x^i, we showed that $\bar{\xi}^i$ in another coordinate system does not have constant components over the space. Let us see how the components $\bar{\xi}^i$ vary when we go from one point to a neighboring one in space along a curve parametrized with a parameter p; to do this we differentiate $\bar{\xi}^i = (\partial\bar{x}^i/\partial x^k)\xi^k$ with respect to p, remembering that the ξ^k's are constant along the curve by assumption. This gives

$$\frac{d\bar{\xi}^i}{dp} = \frac{\partial^2\bar{x}^i}{\partial x^k\,\partial x^l}\,\frac{dx^l}{dp}\,\xi^k$$

But if we want an intrinsic characterization of a constant vector field, we should relate the increments $d\bar{\xi}^i$ to the barred components $\bar{\xi}^i$ themselves. This can be done by rewriting the last formula in a straightforward way:

$$\frac{d\bar{\xi}^i}{dp} = \left(\frac{\partial^2\bar{x}^i}{\partial x^k\,\partial x^l}\,\frac{dx^l}{d\bar{x}^m}\,\frac{d\bar{x}^m}{dp}\,\frac{\partial x^k}{\partial\bar{x}^j}\right)\bar{\xi}^j$$

which may be written

$$\frac{d\bar{\xi}^i}{dp} = \Gamma^i_{mj}\,\frac{d\bar{x}^m}{dp}\,\bar{\xi}^j \qquad (2.1)$$

where

$$\Gamma^i_{mj} = \frac{\partial^2\bar{x}^i}{\partial x^k\,\partial x^l}\,\frac{\partial x^l}{\partial\bar{x}^m}\,\frac{\partial x^k}{\partial\bar{x}^j} \qquad (2.2)$$

We see that constancy of components in one particular coordinate system requires that, in an arbitrary coordinate system, the much more general law (2.1) holds. Specifically, we recognize that the increment of the vector components is a bilinear function of the components ξ^j of the vector and of the displacement dx^m tangent to the curve along which the transplantation takes place.

In order to generalize the above notions, let us now ignore the origin of formula (2.1) and consider only its differential form

$$d\xi^i = \Gamma^i_{mj}\, dx^m\, \xi^j \tag{2.3}$$

The Γ^i_{mj} coefficients are now considered to be any set of given functions of the coordinates and are no longer restricted to have the form (2.2), which was derived from the fact that the vector field considered had constant components in the particular original coordinate system. The vector field ξ^i is considered to be obtained from its value at one given point by the transplantation law (2.3). Equation (2.3) defines a general law for transplantation of the vector ξ^i at the point x into the quantities $\xi^i + d\xi^i$ at the point $x + dx$. It is a law of affine character; that is, it has invariant structure under a linear transformation of the coordinates.

If we now try to make this law of transplantation coordinate-invariant and demand that $\xi^i + d\xi^i$ still be a vector at the point $x + dx$, we shall be forced into certain requirements on the Γ^i_{mj} coefficients; we shall see that these requirements define a transformation law for the Γ^i_{mj} coefficients and therefore allow us to transplant a vector by infinitesimal amounts in a covariant fashion.

Proof. In the unbarred coordinate system we take for the transplantation law

$$\xi^i(x + dx) = \xi^i + d\xi^i = \xi^i + \Gamma^i_{mj}\, dx^m\, \xi^j \tag{2.4}$$

and impose the requirement that the same law hold in the barred coordinate system (that is, that the law be covariant). We further require that $\xi^i(x + dx)$ be a vector; by definition this means

$$\bar{\xi}^j(x + dx) = \xi^i(x + dx)\left(\frac{\partial \bar{x}^j}{\partial x^i}\right)_{x+dx}$$

In this expression the geometric point at which the transformation is carried out is characterized by the markers $x + dx$ (which could also be called $\bar{x} + d\bar{x}$ since the two sets of markers are in one-to-one correspondence). Writing out this expression with the help of the transplantation law (2.4), we have

$$\bar{\xi}^j + \bar{\Gamma}^j_{ms}\, d\bar{x}^m\, \bar{\xi}^s = (\xi^i + \Gamma^i_{ml}\, dx^m\, \xi^l)\left(\frac{\partial \bar{x}^j}{\partial x^i}\right)_{x+dx}$$

Expanding the last factor in a Taylor series and keeping only terms up to the first order, we obtain

$$\left(\frac{\partial \bar{x}^j}{\partial x^i}\right)_{x+dx} = \left(\frac{\partial \bar{x}^j}{\partial x^i}\right)_x + \frac{\partial^2 \bar{x}^j}{\partial x^i\, \partial x^\eta}\, dx^\eta$$

Putting this into the previous equation and relabeling the dummy indices, we find that

$$\bar{\Gamma}^j_{ms}\, d\bar{x}^m\, \bar{\xi}^s = \left(\Gamma^i_{\alpha\beta}\frac{\partial \bar{x}^j}{\partial x^i} + \frac{\partial^2 \bar{x}^j}{\partial x^\beta\, \partial x^\alpha}\right)\xi^\beta\, dx^\alpha$$

But we can also express $dx^\alpha \xi^\beta$ in terms of barred quantities as

$$\xi^\beta\, dx^\alpha = \frac{\partial x^\alpha}{\partial \bar{x}^m}\frac{\partial x^\beta}{\partial \bar{x}^s}\,\bar{\xi}^s\, d\bar{x}^m$$

and obtain

$$\bar{\Gamma}^j_{ms}\, d\bar{x}^m\, \bar{\xi}^s = \left(\frac{\partial \bar{x}^j}{\partial x^i}\frac{\partial x^\alpha}{\partial \bar{x}^m}\frac{\partial x^\beta}{\partial \bar{x}^s}\Gamma^i_{\alpha\beta} + \frac{\partial^2 \bar{x}^j}{\partial x^\alpha\, \partial x^\beta}\frac{\partial x^\alpha}{\partial \bar{x}^m}\frac{\partial x^\beta}{\partial \bar{x}^s}\right)d\bar{x}^m\,\bar{\xi}^s$$

In this equation the coefficients $d\bar{x}^m$ and $\bar{\xi}^s$ are arbitrary; therefore we must have

$$\bar{\Gamma}^j_{ms} = \frac{\partial \bar{x}^j}{\partial x^i}\frac{\partial x^\alpha}{\partial \bar{x}^m}\frac{\partial x^\beta}{\partial \bar{x}^s}\Gamma^i_{\alpha\beta} + \frac{\partial^2 \bar{x}^j}{\partial x^\alpha\, \partial x^\beta}\frac{\partial x^\alpha}{\partial \bar{x}^m}\frac{\partial x^\beta}{\partial \bar{x}^s} \tag{2.5}$$

which is the transformation law of the coefficients $\Gamma^i_{\alpha\beta}$ into which we are forced by the covariance requirement. These coefficients are called *coefficients of affine connection* or simply *connections.* Note that the transformation law (2.5) is *inhomogeneous* in the coefficients $\Gamma^i_{\alpha\beta}$; thus it is fundamentally different from a tensor transformation law and *the connections are not tensors.*

With the above transformation law for $\Gamma^i_{\alpha\beta}$ we have set up a consistent covariant definition of the transplantation of a vector in terms of the increments of its components:

$$d\xi^i = \Gamma^i_{mj}\, dx^m\, \xi^j \tag{2.6}$$

We have now derived necessary conditions for the coefficient set Γ^i_{kl} in order that the transplantation law (2.6) may yield a vector $\xi^i + d\xi^i$ at the point $x^m + dx^m$ if ξ^i is a vector at the point x^m. The next question is how far this transplantation law is sufficient and how it works in the large. For this purpose we assume a field of connections $\Gamma^i_{kl}(x^m)$, that is, a coefficient set attached to each point of the finite part of space considered and which obeys the connection transformation law (2.5). Then we set up the differential-equation system for a vector field in the parameter p:

$$\frac{d\xi^i}{dp} = \Gamma^i_{kl} \frac{dx^k}{dp} \xi^l(p) \tag{2.6'}$$

which allows us to compute the vectors $\xi^i(p)$ along the given curve $x^i(p)$ from the known value $\xi^i(0)$ at the initial point $p = 0$. But we still have to prove that the n-tuples $\xi^i(p)$ thus calculated transform indeed like vectors under a change of coordinates.

To do this we consider the barred coordinates $\bar{x}^i(p)$ and the quantities $\bar{\xi}^i(p)$ defined by the corresponding system (2.6′) in barred terms. We suppose that the equations

$$\bar{\xi}^i(p) - \frac{\partial \bar{x}^i}{\partial x^m} \xi^m(p) = 0$$

hold for $p = 0$ since we assumed at least that $\xi^i(0)$ is a vector. By use of (2.6′) and the analogous differential equation for $\bar{\xi}^i(p)$, we can calculate for all values of p the derivative

$$\frac{d}{dp}\left[\bar{\xi}^i(p) - \frac{\partial \bar{x}^i}{\partial x^m}\xi^m(p)\right] = \bar{\Gamma}^i_{kl}\frac{d\bar{x}^k}{dp}\bar{\xi}^l - \frac{\partial^2 \bar{x}^i}{\partial x^r\,\partial x^s}\frac{dx^r}{dp}\xi^s - \Gamma^m_{rs}\frac{\partial \bar{x}^i}{\partial x^m}\frac{dx^r}{dp}\xi^s$$

If we now use the transformation law (2.5) which we take as hypothesis, we obtain after easy rearrangement

$$\frac{d}{dp}\left[\bar{\xi}^i(p) - \frac{\partial \bar{x}^i}{\partial x^m}\xi^m(p)\right] = \bar{\Gamma}^i_{kl}\frac{d\bar{x}^k}{dp}\left(\bar{\xi}^l - \frac{\partial \bar{x}^l}{\partial x^\alpha}\xi^\alpha\right)$$

The validity of the vector transformation law of $\xi^i(p)$ along the entire curve follows from the uniqueness theorem for this linear homogeneous differential system which admits the solution $\bar{\xi}^i(p) - \frac{\partial \bar{x}^i}{\partial x^m}\xi^m(p) = 0$ as initial value for $p = 0$.

There are several interesting consequences of the transformation law (2.5):

1. If we restrict ourselves to linear transformations of coordinates, the term $\partial^2\bar{x}^j/\partial x^\alpha\,\partial x^\beta$ vanishes and the set of connections Γ^i_{mj} transforms like a tensor. If now some transplanted vector field ξ^i has constant components in a particular coordinate system, it is clear that the terms $\Gamma^i_{mj}\xi^j$ will be zero in that system, as one can see from (2.6). Furthermore, since the Γ transform like a tensor under linear transformations and ξ^j is a vector, these expressions are zero in all coordinate systems related to the original by a linear transformation; it follows that ξ^i has constant components in all such coordinate systems. In this special case constancy of components is therefore an acceptable criterion for a constant vector field.

2. If two fields of connections are given, say, Γ^i_{kl} and $\tilde{\Gamma}^i_{kl}$, their difference is a tensor. Indeed, since the inhomogeneous terms in the transformation law (2.5) are independent of the individual connections, they cancel under subtraction, and $\Gamma^i_{kl} - \tilde{\Gamma}^i_{kl}$ transforms like a tensor. This observation is of particular interest when one varies the connections over a given manifold. The variation of the connection is then a tensor.

3. We can, furthermore, make certain axiomatic deductions from the general form of (2.5). Historically, the connections were introduced in classical differential geometry; their role in transplanting vector fields was first clearly brought out by Levi-Civita. In all these formulations they appeared to be symmetric in their lower indices as in (2.2). But in our more general development we have built up a consistent transplantation law of a vector without the assumption of any particular symmetry of the Γ coefficients. This was pointed out in 1950 by Einstein (Einstein, 1955) and considered independently by Schrödinger (Schrödinger, 1950). Suppose, therefore, that we are given a set of Γ^i_{ms} coefficients such that $\Gamma^i_{ms} \neq \Gamma^i_{sm}$ in a particular coordinate system; we then say that:

a. The Γ coefficients remain unsymmetric under any change of coordinates.

b. It is impossible to find a coordinate system in which all Γ coefficients are 0 at a point.

The proof of these statements is quite simple. For the first one, suppose that we could find a coordinate transformation in which the Γ^i_{mj} coefficients were symmetric; then (2.5) would give $\bar{\Gamma}^i_{\alpha\beta} = \bar{\Gamma}^i_{\beta\alpha}$ in *any* coordinate system, which contradicts our hypothesis. For the second we proceed as above: if such a coordinate system existed, then, by (2.5), the Γ^i_{ms} coefficients would be symmetric in their lower indices in any coordinate system since the inhomogeneous term in (2.5) is symmetric in α and β; this contradicts the hypothesis and completes the proof.

The use of unsymmetric Γ^i_{ms} coefficients has been considered only in later developments of the theory of general relativity, in the attempt to unite electromagnetic theory and gravitation theory. We shall deal in this book only with the original "classical" form of general relativity theory, which uses symmetric Γ^i_{ms} coefficients. It is indeed only in this case that one can connect the law of transplantation (2.3) with the intuitive idea of transplantation in a Euclidean space. This will be proved in the next theorem, which is sometimes referred to as the axiomatic definition of transplantation of Weyl (Weyl, 1950). In a Euclidean space described by a rectilinear coordinate system, we are used to transplanting vectors by simply keeping their components constant and attaching them to different points; in a Riemann space we expect to be able to do the same thing *locally* (in the neighborhood of a point) if we choose the right type of coordinate system at that point. In such a coordinate system we should have $d\xi^i = 0$ and therefore $\bar{\Gamma}^j_{ms} = 0$. Considering the inhomogeneous character of the transformation (2.5), we should often expect to be able to find such a coordinate system.

In the theory of surfaces in three-space, the original object of tensor analysis, the manifold is imbedded in a Euclidean space which determines the metric in the surface. Here transplantation of a vector might be defined by constancy of components in the extrinsic Cartesian coordinate system of the three-space. Locally, we can always find a coordinate system in the surface which coincides with two of the space coordinates. This particular situation provides motivation and illustration for Weyl's conception of transplantation. Weyl considered only transplantations that belong to connections which are symmetric in their lower indices. He showed that such transplantations could be completely characterized by the statement: At every point of the manifold there exists a local coordinate system in which the fields of the transplanted vectors possess constant components under infinitesimal displacement from that point.

From our more general point of view, Weyl's statement is contained in the following theorem.

Theorem. The necessary and sufficient condition for the existence of a particular local coordinate system in which the components of a vector are not altered by an infinitesimal transplantation according to the law (2.3) is that the coefficients of affine connection be symmetric in their lower indices.

Proof of the Necessary Condition. Suppose that, in a particular coordinate system $\bar{x}^i$, the components of an arbitrary vector ξ^s are unaltered under an infinitesimal transplantation from a given point P. This means that $d\xi^i = 0$ in that coordinate system, and therefore that $\bar{\Gamma}^j_{ms}\, d\bar{x}^m \xi^s = 0$. The product $d\bar{x}^m\, \xi^s$ is arbitrary; therefore $\bar{\Gamma}^j_{ms} = 0$ must be satisfied. Then, from (2.5), the coefficients $\Gamma^i_{\alpha\beta}$ will be symmetric in their lower indices in *any* coordinate system, for the inhomogeneous term and the coefficient of $\Gamma^i_{\alpha\beta}$ in (2.5) are symmetric in α and β.

Proof of the Sufficient Condition. Let us choose the point P as the origin of the coordinate system $x^i = 0$, and let us look for a particular coordinate system $\bar{x}^i$ (as mentioned in the theorem) by setting up the transformation

$$\bar{x}^i = x^i + \tfrac{1}{2}A^i_{jk}x^jx^k \qquad \left(\frac{\partial \bar{x}^i}{\partial x^j}\right)_{x=0} = \delta^i{}_j$$

where the A^i_{jk} coefficients have yet to be specified. From (2.5) the Γ^i_{ms} coefficients transform from unbarred to the above defined barred coordinates according to the equation

$$\bar{\Gamma}^j_{ms} = \Gamma^i_{\alpha\beta}\,\delta^j{}_i\,\delta^\alpha{}_m\,\delta^\beta{}_s + \tfrac{1}{2}(A^j_{\alpha\beta} + A^j_{\beta\alpha})\,\delta^\alpha{}_m\,\delta^\beta{}_s$$

which becomes

$$\bar{\Gamma}^j_{ms} = \Gamma^j_{ms} + \tfrac{1}{2}(A^j_{ms} + A^j_{sm})$$

The quantity $(A^j_{ms} + A^j_{sm})$ is symmetric in s and m. With the hypothesis that Γ^j_{ms} is also symmetric, we can choose the A^j_{ms} coefficients so that $\frac{1}{2}(A^j_{ms} + A^j_{sm}) = -\Gamma^j_{ms}$ and obtain $\bar{\Gamma}^j_{ms} = 0$; therefore $d\xi^i = 0$. Thus determined, the A^j_{ms} coefficients define a coordinate system in which the intuitive notion of transplantation by constancy of components is *locally* applicable.

The coordinate system obtained in the preceding paragraph is referred to as a *geodesic coordinate system* with respect to the connection Γ; it is clearly defined only locally. Furthermore, it is defined only up to a linear transformation, as we shall now show. From (2.5) we see that two coordinate systems x^i and $\bar{x}^j$ in which $\Gamma^i_{\alpha\beta}$ and $\bar{\Gamma}^j_{ms}$ are both zero must be related by

$$\frac{\partial^2 \bar{x}^j}{\partial x^\alpha\, \partial x^\beta}\frac{\partial x^\alpha}{\partial \bar{x}^m}\frac{\partial x^\beta}{\partial \bar{x}^s} = 0$$

Consider these equations as a system of four homogeneous equations labeled by s:

$$\left(\frac{\partial^2 \bar{x}^j}{\partial x^\alpha\, \partial x^\beta}\frac{\partial x^\alpha}{\partial \bar{x}^m}\right)\frac{\partial x^\beta}{\partial \bar{x}^s} = 0$$

with $\dfrac{\partial^2 \bar{x}^j}{\partial x^\alpha\, \partial x^\beta}\dfrac{\partial x^\alpha}{\partial \bar{x}^m}$ as unknowns; their determinant is the Jacobian of the coordinate transformation, which we always assume is different from zero.

Therefore the only solution of the system is

$$\frac{\partial^2 \bar{x}^j}{\partial x^\alpha\,\partial x^\beta}\frac{\partial x^\alpha}{\partial \bar{x}^m} = 0$$

But this again is a system of four homogeneous equations labeled by m, whose only solution is

$$\frac{\partial^2 \bar{x}^j}{\partial x^\alpha\,\partial x^\beta} = 0$$

This shows that x^i and $\bar{x}^j$ are indeed related locally by a linear transformation as stated above. A linear one-to-one coordinate transformation is called an affine transformation. The connections $\Gamma^i_{\alpha\beta}$ are called affine connections because of their linear transformation law (2.5).

As mentioned in our remark in Sec. 1.4, the existence of a geodesic coordinate system at each point in space provides us with a powerful tool in tensor calculus because of the very simple form which the $\Gamma^i_{\alpha\beta}$ coefficients have in such a coordinate system; that is, they are zero at the selected point. In the following chapters we shall always assume that the $\Gamma^i_{\alpha\beta}$ coefficients are symmetric in $\alpha\beta$ and that a geodesic coordinate system therefore exists. It should be kept in mind that the $\Gamma^i_{\alpha\beta}$ coefficients do not transform like tensors, but according to (2.5).

We end this section by mentioning some definitions often used in the literature in connection with the law of vector transplantation. A manifold in which a law of vector transplantation is defined is called "affinely related manifold," or an "affine space," and the Γ coefficients are called the "affine connections." We shall use the name *law of vector transplantation* for the general case treated in this section. Some authors use the name "law of parallel displacement" interchangeably with this, but we shall reserve the latter term for the more specific case of a metric space, which will be treated in the next section.

2.2 Parallel Displacement—Christoffel Symbols

In the previous section we introduced a law of *vector transplantation* on an affinely related manifold with the help of the coefficients of affine connection Γ. No metric properties were required to carry out such a program. In this section we particularize our study to Riemann spaces; we shall impose on the transplantation law previously defined the metric requirement that the scalar product of two vectors be invariant under the transplantation. In particular, the length of a vector will then remain

unchanged under the transplantation, as is the case with rectilinear coordinates in Euclidean space. This will give a unique determination of the Γ^i_{jk} coefficients as functions of the components of the metric tensor g_{ik} and their first derivatives.

Let us consider an infinitesimal displacement along a curve and express the fact that the scalar product of two vectors ξ^i and η^k remains constant as they are transplanted along the curve. The scalar product is $g_{ik}\xi^i\eta^k$, so our condition is

$$\frac{d}{ds}\,(g_{ik}\xi^i\eta^k) = 0$$

where ds is the element of arc along the curve. Expanded, this becomes

$$\frac{\partial g_{ik}}{\partial x^l}\frac{dx^l}{ds}\,\xi^i\eta^k + g_{ik}\frac{d\xi^i}{ds}\,\eta^k + g_{ik}\xi^i\frac{d\eta^k}{ds} = 0$$

Taking into account the transplantation law (2.1) for $d\xi^i/ds$ and $d\eta^k/ds$ and relabeling the dummy indices, we obtain for the coefficients of the arbitrary multinomial $\xi^i\eta^k\,dx^l$ the equation

$$\text{(2.7}a\text{)}\qquad \frac{\partial g_{ik}}{\partial x^l} + g_{rk}\Gamma^r_{il} + g_{ir}\Gamma^r_{lk} = 0$$

From this equation we can obtain a unique determination of Γ^r_{il}. Let us cyclically permute the indices ikl in (2.7a) to obtain the following two additional equations:

$$\text{(2.7}b\text{)}\qquad \frac{\partial g_{kl}}{\partial x^i} + g_{rl}\Gamma^r_{ki} + g_{kr}\Gamma^r_{il} = 0$$

$$\text{(2.7}c\text{)}\qquad \frac{\partial g_{li}}{\partial x^k} + g_{ri}\Gamma^r_{lk} + g_{lr}\Gamma^r_{ki} = 0$$

Using the symmetry of g_{ik} and of Γ^l_{ik}, we obtain, by adding (2.7c) and (2.7b) and subtracting (2.7a),

$$\frac{\partial g_{kl}}{\partial x^i} + \frac{\partial g_{li}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^l} + 2\Gamma^r_{ki}g_{rl} = 0$$

Thus

$$\Gamma^r_{ki} = -\tfrac{1}{2}g^{lr}\left(\frac{\partial g_{kl}}{\partial x^i} + \frac{\partial g_{li}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^l}\right)$$

For convenience in notation let us define

$$[ik,l] = \frac{1}{2}\left(\frac{\partial g_{il}}{\partial x^k} + \frac{\partial g_{kl}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^l}\right) \tag{2.8}$$

which we call the Christoffel symbol of the first kind (Christoffel, 1869), and

$$\left\{ {j \atop i\ k} \right\} = g^{jl}[ik,l] \tag{2.9}$$

which we call the *Christoffel symbol of the second kind.* With this notation the coefficients of connection can be simply written as

$$\Gamma^r_{ik} = -\left\{ {r \atop i\ k} \right\} \tag{2.10}$$

We now possess the unique determination of the coefficients of connection for which the scalar product of two vectors remains constant under the law of vector transplantation. We thus arrive at the *law of parallel displacement* in a metric space:

$$d\xi^i = -\left\{ {i \atop \alpha\ \beta} \right\} dx^\alpha\, \xi^\beta \tag{2.11}$$

The behavior of $\left\{ {i \atop \alpha\ \beta} \right\}$ under a coordinate transformation is the same as that of $-\Gamma^i_{\alpha\beta}$, which is given by (2.5).

At this point it might be appropriate to inject a word of comfort to the physicist. The introduction of complicated new symbols in the midst of tensor formalism which is already loaded with index conventions may seem unpleasant to a physicist, and the whole subject may appear to be hidden behind symbolism. In fact, we shall show here that Christoffel symbols actually occur in mechanics, but rarely appear in explicit form because of simplifications due to the very simple mechanical systems usually considered.

Let us consider the evolution in time of a mechanical system described by generalized coordinates $x^i(t)$, generalized velocities $\dot{x}^i = dx^i/dt$, a kinetic-energy quadratic form $T = \frac{1}{2}g_{ik}\dot{x}^i\dot{x}^k$, and a potential energy $V(x^i)$ which gives rise to a generalized force $F_i = -\partial V/\partial x^i$. As usual in analytical dynamics, we take $T\,dt^2 = ds^2$ to define a metric on the space of the generalized coordinates, which is called *configuration space.* In terms of the Lagrangian, $L = T - V$, we can write down the Lagrange equations of motion

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) = \frac{\partial L}{\partial x^i} \tag{2.12}$$

In explicit form these become

$$g_{ik}\ddot{x}^k + \frac{\partial g_{ik}}{\partial x^l}\,\dot{x}^l\dot{x}^k = \frac{1}{2}\frac{\partial g_{lk}}{\partial x^i}\,\dot{x}^l\dot{x}^k + F_i$$

which we can rewrite as

$$g_{ik}\ddot{x}^k + \frac{1}{2}\left[\frac{\partial g_{ik}}{\partial x^l} + \frac{\partial g_{il}}{\partial x^k} - \frac{\partial g_{lk}}{\partial x^i}\right]\dot{x}^l\dot{x}^k = F_i$$

or multiplying by g^{ik},

$$\ddot{x}^i + \left\{ {i \atop l\ k} \right\} \dot{x}^l\dot{x}^k = F^i \tag{2.13}$$

This shows that the Lagrange equations in the general case are second-order differential equations in which Christoffel symbols occur explicitly. We see that, in the case of force-free motion, the generalized velocity vector $\dot{x}^i$ is displaced parallel to itself along the trajectory $x^i(t)$. The physico-geometric significance of such vector transplantation becomes evident.

2.3 Geodesics in Affine and Riemann Space

Let us look for a definition of a straight line in an affine space. In Euclidean space we can characterize such a line by the property that an arbitrary tangent vector along it remains parallel to itself when displaced along the curve. With the help of the law of vector transplantation introduced in Sec. 2.1, we can use the same characteristic property to define a *generalized straight line* in an affine space; we shall call such a curve a *geodesic.* Consider a curve $x^i(q)$ parametrized by q; if $\xi^i(q)$ is an arbitrary tangent vector to the curve, the equation

$$\frac{d\xi^i}{dq} - \Gamma^i_{\alpha\beta}\frac{dx^\alpha}{dq}\,\xi^\beta = 0$$

expresses the fact that it is displaced "unchanged" along the curve. A particular tangent vector is dx^i/dq; a more general tangent vector is therefore $\lambda(q)(dx^i/dq)$, where $\lambda(q)$ is an arbitrary function of q. The equations which define a geodesic in an affine space become

$$\frac{d}{dq}\left(\lambda(q)\,\frac{dx^i}{dq}\right) = \Gamma^i_{\alpha\beta}\,\frac{dx^\alpha}{dq}\,\lambda(q)\,\frac{dx^\beta}{dq} \tag{2.14}$$

This contains the apparently arbitrary function $\lambda(q)$. However, these equations can be brought into a standard form by the following simple transformation: we multiply both sides by $\lambda(q)$ and introduce a new parameter $p(q)$ defined by $dp = dq/[\lambda(q)]$, in terms of which the above equations take the so-called "normal form"

$$\frac{d^2x^i}{dp^2} - \Gamma^i_{\alpha\beta}\,\frac{dx^\alpha}{dp}\,\frac{dx^\beta}{dp} = 0 \tag{2.15}$$

From the way in which we obtained Eqs. (2.15) we see that, starting from an arbitrary parameter q, we can always find a parameter $p(q)$ in terms of which the equations defining geodesic lines in affine space take the "normal form" (2.15). We therefore take these equations to be a *definition of geodesic lines in an affine space.*

One should note that the "normal form" is not parameter-independent and is only valid when one uses a particular class of parameters p; indeed, if we change p into another parameter $\pi(p)$, Eqs. (2.15) become, using $\dfrac{dx^i}{dp} = \dfrac{dx^i}{d\pi}\dfrac{d\pi}{dp}$,

$$\frac{dx^i}{d\pi}\,\frac{d^2\pi}{dp^2} + \left[\frac{d^2x^i}{d\pi^2} - \Gamma^i_{\alpha\beta}\,\frac{dx^\alpha}{d\pi}\,\frac{dx^\beta}{d\pi}\right]\left(\frac{d\pi}{dp}\right)^2 = 0$$

This is still in the "normal form" only if

$$\frac{dx^i}{d\pi}\,\frac{d^2\pi}{dp^2} = 0$$

Since the tangent vector $dx^i/d\pi$ can always be taken to be nonzero, the above condition becomes $d^2\pi/dp^2 = 0$; the parameters π and p must therefore be *proportional*. This proportionality shows that the parameter p of the normal form is determined up to a constant factor; thus we have, even in a nonmetric affine space, a sort of pseudo-length.

Equations (2.15) are a system of ordinary second-order differential equations. Consequently, the solution of the initial-value problem for the geodesics defined by (2.15) [$x^\alpha(0)$ and $dx^\alpha(0)/dp$ are given at the origin, $p = 0$] is unique. Geometrically, this means that, through a given point and with a given tangent at that point, one and only one geodesic line can be drawn.

In the more particular case of a Riemann space we shall use the law of parallel displacement to define geodesics instead of the general law of vector transplantation; the tangent vector which we displace along the curve must in this case be of constant length along the curve. The only such vector is the unit tangent vector dx^i/ds (or a vector proportional to it), where s is the arc length of the geodesic; for this particular tangent vector

$$g_{ik}\,\frac{dx^i}{ds}\,\frac{dx^k}{ds} = \left(\frac{ds}{ds}\right)^2 = 1$$

Furthermore, the affine connections can be expressed explicitly in terms of Christoffel symbols, and the defining equations for a geodesic in a Riemann space become

$$\frac{d^2x^i}{ds^2} + \begin{Bmatrix} i \\ \alpha\ \beta \end{Bmatrix} \frac{dx^\alpha}{ds}\,\frac{dx^\beta}{ds} = 0 \tag{2.16}$$

Other definitions of geodesics in a Riemann space. A geodesic can also be defined by the requirement that it be the shortest curve between a point P_0 and another point P_1. The arc length of a curve between a point P_0 (characterized by the value $p = p_0$ of the curve parameter) and a point P_1 (at which $p = p_1$) is

$$s = \int_{P_0}^{P_1}\left(g_{ik}\,\frac{dx^i}{dp}\,\frac{dx^k}{dp}\right)^{1/2} dp \tag{2.17}$$

As an alternative definition of a geodesic, we require that the variation of this integral be zero. [Note that we must still show consistency with (2.15).] Thus we characterize a geodesic by the requirement

$$\delta\int_{P_0}^{P_1}\left(g_{ik}\,\frac{dx^i}{dp}\,\frac{dx^k}{dp}\right)^{1/2} dp = 0 \tag{2.18}$$

The condition (2.18) expresses a necessary condition for the shortest path between two given points in space (in fact, the mathematical requirement for a stationary path). We might expect this definition of a geodesic to be parameter-independent; in fact, p may be any parameter describing the curve. The change of p into q merely introduces the derivative dq/dp which cancels out of (2.18) because of the homogeneity in dp; therefore the form of Eq. (2.18) remains unaltered.

The definition (2.18) of a geodesic involves the square root of the purely geometric quantity $T = \frac{1}{2}g_{ik}\frac{dx^i}{dp}\frac{dx^k}{dp}$. It has the great theoretical advantage of yielding a parameter-independent integral, but the presence of the square root leads to cumbersome calculations in most applications. It is therefore of practical value to achieve greater flexibility and to show that a more general variational problem leads to precisely the same extremal curves. Let $F(T)$ be an *arbitrary* monotonic and differentiable function of its argument T, and consider the variational problem

$$\delta \int_{P_0}^{P_1} F(T)\, dp = 0 \tag{2.19}$$

This problem depends, of course, upon the particular choice of the parameter p. We shall assume that p is the arc-length parameter s of the extremal curve. More precisely, we deal with the following variational problem: A curve $x^i(s)$ is parametrized by its arc-length parameter s and is compared with all nearby curves $\tilde{x}^i(s)$ which coincide with it at the endpoints $P_0, x^i(0)$ and $P_1, x^i(l)$. What is the condition that, with this parameter $p = s$, the original curve is stationary in this family of competing curves?

Since $g_{ik} = g_{ik}(x^i)$ and $T = \frac{1}{2}g_{ik}\dot{x}^i\dot{x}^k$ (where $\dot{x}^i \equiv dx^i/dp$), Eq. (2.19) can be written with clearly displayed arguments:

$$\delta \int_{P_0}^{P_1} F[T(x^i,\dot{x}^i)]\, dp = 0$$

This is a typical problem of the calculus of variation, which leads to the Euler-Lagrange equations

$$\frac{d}{dp}\left(\frac{\partial F}{\partial \dot{x}^i}\right) = \frac{\partial F}{\partial x^i} \tag{2.20}$$

These may be written, using the functional form of T, as

$$\frac{d}{dp}\left[F'(T)g_{ik}\frac{dx^k}{dp}\right] = F'(T)\,\frac{1}{2}\frac{\partial g_{lk}}{\partial x^i}\frac{dx^l}{dp}\frac{dx^k}{dp}$$

If we make the particular choice of parameter, p = arc length s of the extremal line, we always have $T = \frac{1}{2}g_{ik}\dot{x}^i\dot{x}^k = \frac{1}{2}(ds/ds)^2 = \frac{1}{2}$ along the solution curve of the variational problem. Therefore $F(T)$ and $F'(T)$ are constant along the extremal curve. Thus we can take $F'(T)$ out of the left-hand-side bracket in these equations and still obtain equations defining the extremal curves of the problem:

$$\frac{d}{ds}\left(g_{ik}\frac{dx^k}{ds}\right) = \frac{1}{2}\frac{\partial g_{lk}}{\partial x^i}\frac{dx^l}{ds}\frac{dx^k}{ds}$$

By definition of the Christoffel symbols, and by the same calculations which led from (2.12) to (2.13), we obtain

$$\frac{d^2x^i}{ds^2} + \left\{ {i \atop l\ k} \right\}\frac{dx^l}{ds}\frac{dx^k}{ds} = 0 \tag{2.21}$$

This is identical with the original definition of a geodesic line in a Riemann space (2.16).

We have shown that the variational problem (2.19) and the differential equation (2.16) are consistent alternative definitions of a geodesic line when the curve parameter is the extremal arc length s. The special case $F(T) = \sqrt{T}$ in (2.18) is, however, parameter-independent, as we noted. So (2.18) is consistent with the differential equations (2.16), with no restriction on the choice of curve parameter.

In conclusion, one should notice that, among the three definitions of geodesics which we introduced, the two based on a variational principle, (2.18) and (2.19), are particularly well adapted to the spirit of general relativity because they do not require any particular specification of a coordinate system; therefore they fit immediately into the coordinate-invariant formulation of physics at which one aims in the theory of general relativity.

Connection with dynamics

1. Equations (2.13),

$$\ddot{x}^i + \left\{ {i \atop k\ l} \right\}\dot{x}^k\dot{x}^l = F^i$$

which we obtained as the normal form of the Lagrange equations, are the classical equations of motion of a particle in configuration space. If there are no external forces, $F^i = 0$, they become identical with (2.16) and define *geodesic lines* in configuration space. But in (2.13) we know that the parameter with respect to which the differentiation is performed must be the time parameter t along the trajectory. On the other hand, in (2.16) the parameter must be the arc length s along the geodesic in configuration space. Therefore, since both are normal parameters, the arc length s must be proportional to the time parameter t: $ds \propto dt$. One should note that this proportionality holds only on the geodesic lines;

along other curves the notions of time of transit and of arc length are different.

The fact that the vector $\dot{x}^i$ does not change its length under parallel displacement can be expressed by the equation

$$T = \tfrac{1}{2} g_{ik}\dot{x}^i\dot{x}^k = \text{const}$$

and is nothing but the principle of conservation of energy. Since

$$s = \int \sqrt{g_{ik}\dot{x}^i\dot{x}^k}\, dt$$

we see that

$$s = \sqrt{2T}\, t$$

which determines the above factor of proportionality.

2. With no external forces present the Lagrangian L and the kinetic energy T are equal, $L = T$. Taking the function $F(T)$ (defined above) as T itself, the trajectories of free particles, which are geodesics in configuration space, satisfy the variational problem

$$\delta \int_{P_0}^{P_1} T\, ds = 0 \tag{2.22}$$

Along a geodesic this is the same as

$$\delta \int_{t_0}^{t_1} T\, dt = 0$$

In this last form we recognize Hamilton's principle. The advantage of Hamilton's principle over Eq. (2.18) in defining geodesic lines lies in the fact that the integrand in Hamilton's principle is always defined whereas the presence of the square root in Eq. (2.18) does not allow us to define geodesics when $T \leq 0$. Since the two definitions are equivalent whenever (2.18) has a meaning, we shall adopt Hamilton's principle as a more general definition of geodesics; this allows in particular the definition of *null* geodesic lines, which will be of importance in Chap. 6, when we consider the trajectories of light rays.

It was noticed by Gauss that, for every given force-free dynamical trajectory, one can find a particular coordinate system in which the equations of motion (2.13) reduce to the Newtonian form $\ddot{x}^i = 0$. The proof is immediate, as follows.

We can choose a coordinate system in which the geodesic line described is the x^1 line, i.e., the trajectory is the curve $x^i = 0$ for $i \neq 1$. Furthermore, we may measure x^1 as the arc length along the trajectory. Thus $\dot{x}^1 = \text{const}$ since we know that arc length and time are proportional along the trajectory. Inserting these values into (2.13) we see that $\left\{ {i \atop 1\ 1} \right\} = 0$ along the entire trajectory and that $\ddot{x}^i = 0$ for all components. It is particularly interesting that the Christoffel symbols $\left\{ {i \atop 1\ 1} \right\}$ can be reduced to zero in the large along the entire trajectory. We shall utilize this fact later to introduce a so-called Gaussian coordinate system in which a normal form of the metric tensor is obtained in the large.

Lagrange's equations in the light of general relativity theory. Suppose one forgets about the physical origin of the generalized coordinates x^i and sees the equations of motion written in the form

$$\ddot{x}^i + \left\{ {i \atop k\ l} \right\} \dot{x}^k\dot{x}^l = F^i \tag{2.23}$$

When thinking in terms of Newtonian mechanics, one would like to see these equations take the form $\ddot{x}^i = \tilde{F}^i$, where $\tilde{F}^i$ represents external forces according to Newton's law. To be able to make this identification one considers the quantities

$$-\left\{ {i \atop k\ l} \right\} \dot{x}^k\dot{x}^l$$

as representing fictitious forces (such as centrifugal and Coriolis forces); these visibly depend on the coordinate system used (through the Christoffel symbols) and are often said to appear because one uses the "wrong kind" of coordinate system, for instance, a system attached to a rotating disk. A "right kind" of coordinate system is of course one in which these fictitious forces simply do not appear. However, one can use the alternative approach of treating all forces equally, be they external, fictitious, or due to a constraint, and accordingly write

$$\ddot{x}^i = \tilde{F}^i \qquad \tilde{F}^i = F^i - \left\{ {i \atop k\ l} \right\} \dot{x}^k\dot{x}^l \tag{2.24}$$

Clearly, the combined force depends very much on the coordinate system.

Obviously, such a viewpoint is not in the spirit of general relativity theory where all kinds of coordinate systems are considered equivalent. From the viewpoint of general relativity, one would instead like to reduce the equations of motion as much as possible to the *geometry* of the configuration space. That is, instead of explaining away wrong geometries by fictitious forces, we should like to explain away forces by proper choices of geometry. This will be possible at least in the case of gravitational forces. The easiest way to do this is to postulate that the gravitational forces F^i can be made to disappear from the above equations of motion by incorporating them into the geometric term $\begin{Bmatrix} i \\ k\ l \end{Bmatrix} \dot{x}^k \dot{x}^l$ just like a fictitious force. This approach is motivated by the fact that gravitational and fictitious forces both act on material bodies in the same way; they communicate an acceleration $\ddot{x}^i$ which is *independent of the body's mass.* (This is not the case for other types of forces; for instance, the acceleration communicated to a body by a spring is inversely proportional to the mass of the body.) The above property is the basis of the *principle of equivalence,* which states that the effect of a gravitational field can be "reproduced" by describing physics in an appropriately accelerated frame of reference without interior gravitational forces present. In such a frame of reference the generalized coordinates will be some y^j (which can be considered functions of x^i and t), and the kinetic energy of the system will be described by a new function $\bar{T}$. In order to bring in the principle of equivalence and incorporate all gravitational forces in the geometric term, one would like the Lagrange equations in the moving frame (with coordinates y^j and kinetic energy function $\bar{T}$) to take the form

$$\ddot{y}^j = - \begin{Bmatrix} j \\ k\ l \end{Bmatrix} \dot{y}^k \dot{y}^l \tag{2.25}$$

which are the equations of configuration-space geodesics in the moving frame.

Unfortunately, we can easily show that such an attempt to incorporate the principle of equivalence cannot succeed within the framework of classical mechanics: consider the concrete case of a particle moving under the influence of gravity along a three-dimensional trajectory described by $y^j(t)$ ($j = 1, 2, 3$) in a moving frame of reference. If Eqs. (2.25) were valid, the acceleration $\ddot{y}^j$ of the particle in that frame of reference would depend quadratically on the velocity $\dot{y}^j$ of the particle; doubling the velocity of a particle submitted to a gravitational field would therefore quadruple its acceleration. We know from experience that the movement of a particle in a gravitational field does not obey such a law in any frame of reference. Equations (2.25) are therefore unacceptable to describe the effects of gravitational forces, and *it is impossible to have gravitational forces take the same mathematical form as fictitious forces within the framework of classical analytical mechanics.*

We shall be able to formulate the principle of equivalence in mathematical terms only when we consider *Euler-Lagrange equations in a four-*

dimensional space which includes time as an ordinary coordinate; in order to consider the solution of gravitational problems by a purely geometrical treatment, it will be necessary to make use of the concepts of special relativity theory and the Lorentz metric. In fact, in a four-dimensional space for which the zeroth coordinate is ct (time multiplied by the speed of light), Eqs. (2.25) are acceptable. When the velocities involved in the problem are small compared with the speed of light, we have

$$\dot{y}^0 = \frac{d}{dt}(ct) = c$$

$$\dot{y}^k \ll c \qquad k = 1, 2, 3$$

Thus Eqs. (2.25) reduce in lowest order to

$$\ddot{y}^j = - \begin{Bmatrix} j \\ 0\ 0 \end{Bmatrix} c^2 \qquad j = 1, 2, 3 \tag{2.26}$$

The terms quadratic in velocity, which prevented us from making any progress in a three-dimensional framework, do not appear in these lower-order equations in a four-dimensional framework. The only term which survives is the constant c^2. Furthermore, we see that the Christoffel symbols (which are geometric entities) here play the role of forces. These considerations will be taken up in greater detail in Chap. 4.

2.4 Gaussian Coordinates

By letting a family of geodesics play a particular role among the coordinate lines, Gauss has shown the existence of a very useful coordinate system, which we shall now describe. Let us restrict ourselves to the case of a four-dimensional space with a hyperbolic metric, which we defined earlier as a metric of signature $(1, -1, -1, -1)$. Consider a three-dimensional hypersurface S imbedded in this four-dimensional space; we suppose that any vector n normal to S satisfies the inequality

$$n^0 n_0 + (n^1 n_1 + n^2 n_2 + n^3 n_3) > 0 \tag{2.27}$$

which, in the familiar language of special relativity theory, implies that the surface is "oriented in space" (whereas a vector normal to S is "oriented in time").

We introduce in the surface S three coordinates x^{*1}, x^{*2}, x^{*3} which serve to characterize the variable point $P^* \in S$. Through each point

P^* of the three-dimensional surface S we draw the geodesic which is orthogonal to S at P^*. These geodesics will form a field of nonintersecting curves in some finite neighborhood M of S such that, through each point P of M, there will pass exactly one of the geodesics constructed. We introduce now, in the entire four-dimensional domain M, coordinates as follows: Given P, we consider the geodesic of the field passing through P and its original point $P^* \varepsilon S$. We define the coordinates x^i of P in

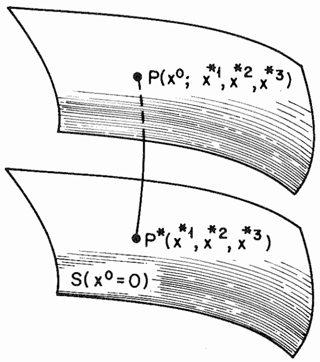


Fig. 2.1

terms of the arc length P^*P of the geodesic and of the coordinates x^{*i} of P^*:

$$x^0 = \text{arc length } P^*P \text{ along the geodesic}$$

$$x^1 = x^{*1}$$

$$x^2 = x^{*2}$$

$$x^3 = x^{*3}$$

In this manner, the three coordinates x^1, x^2, x^3 remain constant along any geodesic perpendicular to S; it follows that, along such a geodesic,

$$ds^2 = (dx^0)^2 \qquad g_{00} = 1$$

Next let us express the conditions resulting from the orthogonality of the x^0 (geodesic) lines to the hypersurface S; any vector $(0,a,b,c)$ in S must be orthogonal to the vector $(1,0,0,0)$ tangent to the x^0 line at the

same point, which requires that

$$g_{01} = g_{02} = g_{03} = 0$$

on the hypersurface S. Thus we see that, on S, the line element has the form

$$(dx^0)^2 + g_{ik}\, dx^i\, dx^k$$

Let us now attempt to show that the above form of the line element also holds outside the hypersurface S in the coordinate system we have constructed above. From the equations of a geodesic line (2.16),

$$\frac{d^2x^\alpha}{ds^2} + \begin{Bmatrix} \alpha \\ \mu \;\; \nu \end{Bmatrix} \frac{dx^\mu}{ds}\frac{ds^\nu}{ds} = 0$$

applied to the present case, for which x^0 alone varies along a geodesic and $ds = dx^0$, we deduce

$$\begin{Bmatrix} i \\ 0 \;\; 0 \end{Bmatrix} = 0 \qquad i = 1, 2, 3$$

and therefore

$$[00,i] = 0$$

Explicitly, writing out this last condition, we obtain

$$2\frac{\partial g_{0i}}{\partial x^0} - \frac{\partial g_{00}}{\partial x^i} = 0$$

Since $g_{00} = 1$ on *any* geodesic, the second term in the above is identically zero, so we have

$$\frac{\partial g_{0i}}{dx^0} = 0$$

along the x^0 lines. Therefore the elements g_{01}, g_{02}, g_{03} of the metric tensor, which are zero on S, remain zero in the entire domain M determined by the field of geodesics. In such a domain, the metric of the four-dimensional space expressed in Gaussian coordinates takes everywhere the form

$$ds^2 = (dx^0)^2 + g_{ij}(x^0,x^1,x^2,x^3)\, dx^i\, dx^j \qquad i, j = 1, 2, 3 \tag{2.28}$$

Any hypersurface ($x_0 =$ const) is therefore orthogonal to the geodesic x^0 coordinate lines.

Furthermore, the lengths of the segments of any two geodesic lines between hypersurfaces S_1 and S_2 are clearly equal; the lengths of these segments correspond to the unique time interval Δx^0 between S_1 and S_2. Thus the surfaces $x^0 =$ const are equidistant level surfaces in terms of our geodesics; they correspond to parallel planes in Euclidean geometry.

The Gaussian coordinates introduced here will be very useful in later chapters because of the separation which they perform between a time coordinate x^0 valid everywhere in three-space and the metrical description of three-space itself. They allow us to connect the abstract four-dimensional framework of general relativity theory with the classical intuitive point of view which regards events as occurring in three-space and describes the development of these events in terms of a universal time parameter. (In Sec. 8.3 there is further discussion of the relation of Gaussian coordinates to the Einstein equations.)

Exercises

2.1 How many algebraically independent Christoffel symbols are there in two, three, and four dimensions? In n dimensions?

2.2 What are the Christoffel symbols in two-dimensional Euclidean space with orthogonal axes? What if the axes are canted, as mentioned in Sec. 1.9?

2.3 What are the Christoffel symbols for the surface of a unit sphere? (Use the results of Exercise 1.6.) Show that the equator and longitude lines on the surface of the earth are geodesics but that latitude lines are not.

2.4 Consider, in spherical coordinates r, θ, and φ, a diagonal metric with diagonal components $f(r)$, r^2, and $r^2 \sin^2 \theta$. Calculate the Christoffel symbols.

2.5 What are the geodesic equations for the metric of Exercise 2.4? Is the ray $\theta =$ const, $\varphi =$ const a geodesic?

2.6 Consider a vector of small coordinate displacements $(\Delta r, \Delta\theta)$ attached to the point (r,θ) in polar coordinates. Parallel-displace it to a nearby point using the law (2.6). Check diagramatically that the displaced vector and the original vector are indeed parallel in the usual Euclidean sense.

2.7 Prove the theorem $g_{il|k} = [ki,l] + [kl,i]$.

Problems

2.1 Use matrix theory to show that by a real coordinate transformation any metric can be brought into diagonal form at a given point. Moreover show that the diagonal elements can be made equal to $+1$, -1, or 0 and that the number of $+1$, -1, and 0 elements is independent of the manner in which the transformation is achieved. This diagonal matrix is called the *Cayley-Sylvester canonical form* of the matrix, and the set of diagonal elements is an alternative definition of the signature [see Eq. (1.2)]. It is always assumed to be $(+1,-1,-1,-1)$ in relativity theory, and the coordinate system defined by the transformation is termed a *tangent Lorentz space;* see Sec. 5.6 for more details.

2.2 The theorem in Sec. 2.1 guarantees that there exists a coordinate system in which the coefficients of affine connection vanish if they are symmetric in any system. Extend that theorem to show that the coordinate system may simultaneously be chosen to be a tangent Lorentz space by considering a transformation $\bar{x}^i = B^i_j x^j + \frac{1}{2} A^i_{jk} x^j x^k$, where the A^i_{jk} and B^i_j are suitably chosen constants.

Bibliography

Cartan, E. (1946): "Leçons sur la géométrie des espaces de Riemann," 2d ed., Paris
Christoffel, E. B. (1869): Über die Transformation der homogenen Differentialausdrücke zweiten Grades, "Journal für die reine und angewandte Mathematik (Crelle)," Vol. 70, pp. 46–70.
Einstein, A. (1955): "Meaning of Relativity," 5th ed., Princeton, N.J.
Eisenhart, L. P. (1949): "Riemannian Geometry," 2d ed., Princeton, N.J.
Kreyszig, E. (1959): "Differential Geometry," Toronto.
Lichnerowicz, A. (1955): "Théories relativistes de la gravitation et de l'électro-magnétisme," Paris.
Rainich, G. Y. (1950): "Mathematics of Relativity," New York–London.
Raschewski, P. K. (1959): "Riemannsche Geometrie und Tensoranalysis," Berlin. (Translation from the Russian, Moscow, 1953).
Schouten, J. A., and D. J. Struik (1935–1938): "Einführung in die neueren Methoden der Differentialgeometrie," 2 vols., Groningen.
Schrödinger, E. (1950): "Space-Time Structure," London.
Weatherburn, C. E. (1938): "Riemannian Geometry and the Tensor Calculus," London.
Weyl, H. (1950): "Space, Time, Matter," New York.
Whittaker, E. T. (1937): "Analytical Dynamics," 4th ed., London.
Willmore, T. J. (1959): "An Introduction to Differential Geometry," Oxford.

CHAPTER

3

Tensor Analysis

We return in this chapter to the problem of comparing tensors attached to neighboring points of space. To do this we shall first define intrinsically a differentiation process for a vector field.

3.1 Covariant Differentiation

Let us study how a contravariant vector field $\xi^i(x^j)$ varies when one goes from a point x^j to a neighboring point $x^j + dx^j$ in an affine space. To do this we shall compare at the point $x^j + dx^j$ the value $\xi^i(x^j + dx^j)$ of the vector field with the vector $\xi^{i*}(x^j + dx^j)$ obtained from $\xi^i(x^j)$ by the law of vector transplantation along the infinitesimal vector dx^j (Fig. 3.1). We thus form, at the point $x^j + dx^j$, the vector difference

$$\xi^i(x^j + dx^j) - \xi^{i*}(x^j + dx^j) \tag{3.1}$$

By the use of a Taylor expansion, $\xi^i(x^j + dx^j)$ can be written as

$$\xi^i(x^j + dx^j) = \xi^i(x^j) + \frac{\partial \xi^i}{\partial x^k}\, dx^k + O(dx^k)^2$$

From the law of vector transplantation, the displaced vector $\xi^{i*}(x^j + dx^j)$ is equal to

$$\xi^{i*}(x^j + dx^j) = \xi^i(x^j) + \Gamma^i_{kl}\xi^l\, dx^k$$

Therefore the vector difference (3.1) becomes

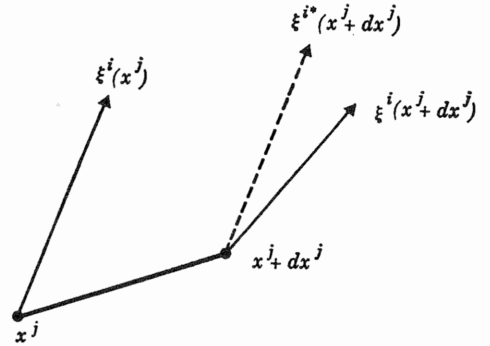
$$\xi^i(x^j + dx^j) - \xi^{i*}(x^j + dx^j) = \left[\frac{\partial \xi^i}{\partial x^k} - \Gamma^i_{kl}\xi^l\right] dx^k + O(dx^k)^2 \tag{3.2}$$

This expression has the form of the beginning of a series expansion and suggests interpreting the quantity

$$\left[\frac{\partial \xi^i}{\partial x^k} - \Gamma^i_{kl}\, \xi^l\right] \tag{3.3}$$

which is independent of the displacement dx^j, as a sort of "first derivative" of the field $\xi^i(x^j)$. Furthermore, formula (3.2) is valid for an *arbitrary* infinitesimal displacement dx. The left side of (3.2) is a vector (the difference of two vectors), and on the right side dx^k is an arbitrary vector. Thus, neglecting terms of higher order than the first in dx^k,

Fig. 3.1



we see by the quotient theorem that the quantity (3.3) is a tensor. We shall call it the *covariant derivative* of the contravariant vector field ξ^i.

One should notice that the derivatives $\partial \xi^i/\partial x^k$ taken with respect to arbitrary coordinates do not themselves form a tensor, as is evident from their transformation properties; neither do the coefficients of connection, as we already know. But the combination (3.3) is, as we have proved, a tensor. At this point it is advantageous for simplicity to introduce certain notation conventions. To indicate *ordinary differentiation* of a tensor component $T^\alpha{}_{\beta\gamma}$ with respect to x^k, we introduce the notation

$$T^\alpha{}_{\beta\gamma|k} = \frac{\partial T^\alpha{}_{\beta\gamma}}{\partial x^k}$$

The bar in front of the index k indicates that this particular index is an index of differentiation, while the others retain their usual meaning. To represent the *covariant derivative* of a vector, we introduce the notation

$$\xi^i_{\|k} = \frac{\partial \xi^i}{\partial x^k} - \Gamma^i_{kl}\xi^l$$

which can also be written in terms of the previous convention as

$$\xi^i_{\|k} = \xi^i_{|k} - \Gamma^i_{kl}\xi^l \tag{3.4}$$

The name covariant derivative given to this tensor is justified for the following reasons:

1. It appears already as playing the role of a first derivative in the Taylor series (3.2), where it first occurs.
2. It reduces to the usual derivative $\xi^i_{|k}$ in a geodesic coordinate system in which the connections Γ^i_{kl} vanish.

If $\xi^i_{\|k}$ were zero over a finite region of space, one could then build up a vector field by vector transplantation, starting with one given vector at one particular point (x^i). Then we should have

$$\xi^i(x^j + dx^j) = \xi^{i*}(x^j + dx^j)$$

over a finite region of space, and ξ^i would obey the law of vector transplantation. Such a vector field could then be called a generalized *constant vector field* since it would have a zero covariant derivative. However, one must be careful to note that the condition

$$\xi^i_{\|k} = \xi^i_{|k} - \Gamma^i_{kl}\xi^l = 0$$

can hold in a finite region of space *only if certain integrability conditions on the connections are satisfied.* These conditions will put restrictions on the type of spaces in which constant vector fields can be defined. Indeed, we shall show in Chap. 5 that a constant vector field can be defined *only* in a pseudo-Euclidean space. It is for this reason that we never attempted to compare vectors which are a finite distance apart by using the vector-transplantation law to carry one vector to the point of attachment of the other vector. We shall show in Chap. 5 that such a procedure depends in general on the path along which the transplantation is performed.

The concept of covariant derivative can be specialized in an obvious way to the case of Riemann spaces by replacing the concept of vector transplantation by that of parallel displacement and by expressing the affine connections in terms of Christoffel symbols. The covariant derivative of a vector field in a Riemann space is then defined by analogy with (3.4) to be

$$\xi^i_{\|k} = \xi^i_{|k} + \begin{Bmatrix} i \\ k\ l \end{Bmatrix} \xi^l \tag{3.5}$$

In the rest of this book we shall always deal with a Riemann space because one uses such a space in the theory of general relativity, but most concepts which will be introduced in this chapter can easily be generalized to an affine space.

We shall now give a direct proof that $\xi^i_{\|k}$ is a tensor, without recourse to series expansions. Given a vector field $\xi^i(x^\alpha)$ in a Riemann space, let us consider an arbitrary curve $x^\alpha(s)$ and define a vector field $\eta^i(s)$ at each point of the curve by the equations $\dfrac{d\eta^i}{ds} + \begin{Bmatrix} i \\ k\ l \end{Bmatrix} \dfrac{dx^l}{ds}\eta^k = 0$. These are ordinary first-order differential equations, and the $\eta^i(s)$ are defined along the curve once initial values have been chosen. The quantity

$$P(s) = g_{ik}\xi^i\eta^k$$

defined for each point of the curve is clearly a scalar; therefore

$$\frac{dP(s)}{ds} = P'(s)$$

is also a scalar. Let us write it out:

$$P'(s) = g_{ik|l}\frac{dx^l}{ds}\xi^i\eta^k + g_{ik}\xi^i_{|l}\frac{dx^l}{ds}\eta^k + g_{ik}\,\xi^i\eta^k_{|l}\frac{dx^l}{ds}$$

Using the defining equation for η^i and relabeling the dummy indices, we obtain

$$P'(s) = g_{ir|l}\,\xi^i\frac{dx^l}{ds}\eta^r + g_{ir}\xi^i_{|l}\frac{dx^l}{ds}\eta^r - g_{ik}\,\xi^i \begin{Bmatrix} k \\ r\ l \end{Bmatrix} \frac{dx^l}{ds}\eta^r$$

or

$$P'(s) = \left[\xi^i\left(g_{ir|l} - g_{ik}\begin{Bmatrix} k \\ r\ l \end{Bmatrix}\right) + g_{ir}\xi^i_{|l}\right]\frac{dx^l}{ds}\eta^r = T_{rl}\frac{dx^l}{ds}\eta^r$$

where T_{rl} represents the bracket. The left-hand side of this equation is a scalar, and dx^l/ds and η^r are arbitrary vectors at x^i; therefore, by the quotient theorem, the quantity T_{rl} *is a tensor*. The form of T_{rl} can be simplified using only the definitions of the Christoffel symbols of the first and second kind, (2.8) and (2.9). One obtains immediately

$$g_{ir|l} - g_{ik}\begin{Bmatrix} k \\ r\ l \end{Bmatrix} = g_{ir|l} - [rl,i] = [il,r] \tag{3.6}$$

Thus we may write T_{rl} as

$$T_{rl} = \xi^i[il,r] + g_{ir}\xi^i_{|l}$$

Multiplication and contraction with g^{rs} then gives the result

$$T^s{}_l = \xi^s_{|l} + \begin{Bmatrix} s \\ i\ l \end{Bmatrix}\xi^i$$

Since T_{rl} is a tensor, $T^s{}_l$ is also a tensor. Thus we have proved in a formal way that $\xi^s_{\|l} = T^s{}_l$ is indeed a tensor. However, our first derivation displays more clearly the significance of the term as a vector derivative under parallel displacement.

Covariant derivative of a covariant vector field. In the foregoing paragraphs we defined the covariant derivative of a contravariant vector field. Let us define an analogous operation for a covariant vector field $\eta_i(x^\alpha)$. We shall impose on the covariant derivative the defining conditions that it be a tensor and reduce to the ordinary derivative in a geodesic coordinate system, which we defined in Chap. 2. Given the covariant vector field $\eta_i(x^\alpha)$, consider an arbitrary contravariant vector field $\xi^i(x^\alpha)$ and form the scalar $\phi(x^\alpha) = \xi^i(x^\alpha)\eta_i(x^\alpha)$. We can write the gradient of the scalar ϕ, which we know is a covariant vector, as

$$v_l = \phi_{|l} = (\xi^i\eta_i)_{|l} = \xi^i_{|l}\eta_i + \xi^i\eta_{i|l}$$

Note that v_l is a *vector field*, which has been created from two other vector fields without the use of Christoffel symbols. By definition, the covariant derivative of ξ^i is

$$\xi^i_{\|l} = \xi^i_{|l} + \begin{Bmatrix} i \\ m\ l \end{Bmatrix}\xi^m$$

Using this tensor, let us form the covariant vector $\eta_i\xi^i_{\|l} = w_l$ and consider the vector difference $s_l = v_l - w_l$. We now have the vector

$$s_l = v_l - w_l = \eta_{i|l}\xi^i - \begin{Bmatrix} i \\ m\ l \end{Bmatrix}\xi^m\eta_i$$

By changing the first dummy index i into m, we may write this as

$$s_l = \xi^m\left(\eta_{m|l} - \begin{Bmatrix} r \\ m\ l \end{Bmatrix}\eta_r\right)$$

Since ξ^m is an arbitrary contravariant vector and s_l is (as we have constructed it) a covariant vector, we may use the quotient theorem to deduce that $\eta_{m|l} - \left\{ {r \atop m\ l} \right\} \eta_r$ is a tensor. Furthermore, since it clearly reduces to the ordinary derivative of η_m in a geodesic coordinate system, we may call it the covariant derivative of the covariant vector η_i. In analogy to the covariant derivative notation of (3.5) for a contravariant vector, we shall denote it by

$$\eta_{m\|l} = \eta_{m|l} - \left\{ {r \atop m\ l} \right\} \eta_r \tag{3.7}$$

Differentiation of a vector product and differentiation of tensors. Consider two arbitrary vector fields ξ^i and η^k and the tensor $\xi^i\eta^k$. We shall define the covariant derivative of this tensor $(\xi^i\eta^k)_{\|l}$ to meet the following requirement: In a geodesic coordinate system this quantity is to be identical with $(\xi^i\eta^k)_{|l} = \xi^i\eta^k_{|l} + \xi^i_{|l}\eta^k$. We know that, in a geodesic coordinate system, $\xi^i_{|l} = \xi^i_{\|l}$, and therefore

$$(\xi^i\eta^k)_{\|l} = \xi^i\eta^k_{\|l} + \xi^i_{\|l}\eta^k$$

But in this last form the right side of the expression is written as a tensor. We can therefore take it as a consistent definition of $(\xi^i\eta^k)_{\|l}$ in an arbitrary coordinate system.

We can evidently extend the previous reasoning to any tensor of the form $\xi^i\eta^k\zeta_m$ and obtain

$$(\xi^i\eta^k\zeta_m)_{\|l} = \xi^i\eta^k\zeta_{m\|l} + \xi^i\eta^k_{\|l}\zeta_m + \xi^i_{\|l}\eta^k\zeta_m \tag{3.8}$$

Replacing the covariant derivatives of the vectors by their values

$$\xi^i_{\|l} = \xi^i_{|l} + \left\{ {i \atop l\ r} \right\} \xi^r$$

etc., we obtain

$$(\xi^i\eta^k\zeta_m)_{\|l} = (\xi^i\eta^k\zeta_m)_{|l} + \left\{ {i \atop l\ r} \right\} \xi^r\eta^k\zeta_m + \left\{ {k \atop l\ r} \right\} \xi^i\eta^r\zeta_m - \left\{ {r \atop l\ m} \right\} \xi^i\eta^k\zeta_r \tag{3.9}$$

Note that, whenever a summation occurs over an index of a covariant vector, the summation index appears in the upper position in the Christoffel symbol; summations involving a contravariant vector have the summation index in the lower position in the Christoffel symbol.

Since we know from Sec. 1.5 that any tensor $T^{\alpha\beta}{}_\gamma$ can be written in the form of a sum of multinomials $T^{\alpha\beta}{}_\gamma = \Sigma\xi^\alpha\eta^\beta\zeta_\gamma$, we have also found the rule of differentiation of any tensor. A typical case is

$$T^{\alpha\beta}{}_{\gamma\|l} = T^{\alpha\beta}{}_{\gamma|l} + \left\{ {\alpha \atop \tau\ l} \right\} T^{\tau\beta}{}_\gamma + \left\{ {\beta \atop \tau\ l} \right\} T^{\alpha\tau}{}_\gamma - \left\{ {\tau \atop \gamma\ l} \right\} T^{\alpha\beta}{}_\tau$$

The generalization to any number of indices is evident; one need only be careful to balance indices and remember to use a plus sign for each contravariant index and a minus sign for each covariant index in the general case.

In the remainder of this book, we shall often follow the usage of physicists and speak of the covariant derivative of a tensor. It would be more precise but more verbose to speak of a tensor field.

Other approach to tensor covariant differentiation formula. Instead of deducing the law of differentiation for tensors from the law known for vectors, we could proceed directly as follows: If $T^\alpha{}_{\beta\gamma}$ is a tensor and ξ_α, η^β, ζ^γ are three arbitrary vectors, consider the scalar $T^\alpha{}_{\beta\gamma}\xi_\alpha\eta^\beta\zeta^\gamma$. Its gradient $(T^\alpha{}_{\beta\gamma}\xi_\alpha\eta^\beta\zeta^\gamma)_{|l}$ is a covariant vector which we shall call w_l:

$$w_l = T^\alpha{}_{\beta\gamma|l}\xi_\alpha\eta^\beta\zeta^\gamma + T^\alpha{}_{\beta\gamma}\xi_{\alpha|l}\eta^\beta\zeta^\gamma + T^\alpha{}_{\beta\gamma}\xi_\alpha\eta^\beta_{|l}\zeta^\gamma + T^\alpha{}_{\beta\gamma}\xi_\alpha\eta^\beta\zeta^\gamma_{|l}$$

Next consider the vector v_l:

$$v_l = T^\alpha{}_{\beta\gamma}(\xi_\alpha\eta^\beta\zeta^\gamma)_{\|l}$$

which can be written by (3.8)

$$v_l = T^\alpha{}_{\beta\gamma}\xi_{\alpha\|l}\eta^\beta\zeta^\gamma + T^\alpha{}_{\beta\gamma}\xi_\alpha\eta^\beta_{\|l}\zeta^\gamma + T^\alpha{}_{\beta\gamma}\xi_\alpha\eta^\beta\zeta^\gamma_{\|l}$$

Let us form the vector difference

$$w_l - v_l = T^\alpha{}_{\beta\gamma|l}\xi_\alpha\eta^\beta\zeta^\gamma + T^\alpha{}_{\beta\gamma}(\xi_{\alpha|l} - \xi_{\alpha\|l})\eta^\beta\zeta^\gamma + T^\alpha{}_{\beta\gamma}\xi_\alpha(\eta^\beta_{|l} - \eta^\beta_{\|l})\zeta^\gamma + T^\alpha{}_{\beta\gamma}\xi_\alpha\eta^\beta(\zeta^\gamma_{|l} - \zeta^\gamma_{\|l})$$

We know from (3.7) that

$$\xi_{\alpha|l} - \xi_{\alpha\|l} = \left\{ {r \atop \alpha\ l} \right\} \xi_r$$

and similarly for the differences $\eta^\beta{}_{|l} - \eta^\beta{}_{\|l}$ and $\zeta^\gamma{}_{|l} - \zeta^\gamma{}_{\|l}$. Substitution of these in the above expression for $w_l - v_l$ gives

$$w_l - v_l = T^\alpha{}_{\beta\gamma|l}\xi_\alpha\eta^\beta\zeta^\gamma + T^\alpha{}_{\beta\gamma}\begin{Bmatrix} r \\ \alpha\ l \end{Bmatrix}\eta^\beta\zeta^\gamma\xi_r - T^\alpha{}_{\beta\gamma}\begin{Bmatrix} \beta \\ l\ r \end{Bmatrix}\xi_\alpha\eta^r\zeta^\gamma - T^\alpha{}_{\beta\gamma}\begin{Bmatrix} \gamma \\ l\ r \end{Bmatrix}\xi_\alpha\eta^\beta\zeta^r$$

Relabeling the dummy indices, we have

$$w_l - v_l = \left[T^\alpha{}_{\beta\gamma|l} + \begin{Bmatrix} \alpha \\ l\ s \end{Bmatrix} T^s{}_{\beta\gamma} - \begin{Bmatrix} s \\ l\ \beta \end{Bmatrix} T^\alpha{}_{s\gamma} - \begin{Bmatrix} s \\ l\ \gamma \end{Bmatrix} T^\alpha{}_{\beta s}\right]\xi_\alpha\eta^\beta\zeta^\gamma$$

Since ξ_α, η^β, ζ^γ are arbitrary vectors and $w_l - v_l$ is a vector, we may use the quotient theorem to infer that the quantity in brackets is a tensor. It reduces to $T^\alpha{}_{\beta\gamma|l}$ in a geodesic coordinate system. We therefore call it the covariant derivative of the tensor; as expected, this definition coincides with (3.9) obtained by our previous method.

$$T^\alpha{}_{\beta\gamma\|l} = T^\alpha{}_{\beta\gamma|l} + \begin{Bmatrix} \alpha \\ l\ s \end{Bmatrix} T^s{}_{\beta\gamma} - \begin{Bmatrix} s \\ l\ \beta \end{Bmatrix} T^\alpha{}_{s\gamma} - \begin{Bmatrix} s \\ l\ \gamma \end{Bmatrix} T^\alpha{}_{\beta s} \tag{3.10}$$

Properties of covariant differentiation. Let us now consider the derivative of the product of two tensors,

$$(T^\alpha{}_{\beta\gamma}S^\epsilon{}_{\sigma\nu})_{\|l}$$

We may ask whether the usual formula for differentiation of a product remains valid:

$$(T^\alpha{}_{\beta\gamma}S^\epsilon{}_{\sigma\nu})_{\|l} = T^\alpha{}_{\beta\gamma\|l}S^\epsilon{}_{\sigma\nu} + T^\alpha{}_{\beta\gamma}S^\epsilon{}_{\sigma\nu\|l}$$

The validity of this formula could be inferred directly from formula (3.8). However, it can be shown to be correct more directly by noting that the two sides of the equation are tensors. We know that they are equal in one particular coordinate system—the geodesic system in which covariant differentiation is equivalent to ordinary differentiation. Thus the equation is a tensor equality and is true in all systems.

We can now prove an important theorem due to Ricci: *The covariant derivative of the g_{ik} tensor is identically zero.* To prove this we first make use of (3.6) to obtain the ordinary derivative

$$g_{ir|l} = [il,r] + g_{ik}\begin{Bmatrix} k \\ r\ l \end{Bmatrix}$$

Thus the covariant derivative is

$$\begin{aligned} g_{ir\|l} &= [il,r] + g_{ik}\begin{Bmatrix} k \\ r\ l \end{Bmatrix} - \begin{Bmatrix} k \\ r\ l \end{Bmatrix} g_{ik} - \begin{Bmatrix} k \\ i\ l \end{Bmatrix} g_{kr} \\ &= [il,r] - g_{kr}\begin{Bmatrix} k \\ i\ l \end{Bmatrix} \end{aligned}$$

Using the definition of the Christoffel symbol $\begin{Bmatrix} k \\ i\ l \end{Bmatrix}$ in (2.9), we see that this is identically zero. Thus the metric tensor which characterizes the Riemann space is a constant in the absolute sense; i.e., it has a zero covariant derivative. From this property it is obvious that the operations of raising or lowering indices commute with covariant differentiation. For instance,

$$(\xi_i)_{\|l} = (g_{ik}\xi^k)_{\|l} = g_{ik\|l}\xi^k + g_{ik}\xi^k{}_{\|l} = g_{ik}(\xi^k{}_{\|l})$$

since $g_{ik\|l} = 0$.

We conclude this section with the important remark that the operation of covariant differentiation does not possess in general the commutative property of ordinary differentiation.

$$\xi^\alpha{}_{\|\mu\|\nu} \neq \xi^\alpha{}_{\|\nu\|\mu} \qquad \text{but} \qquad \xi^\alpha{}_{|\mu|\nu} = \xi^\alpha{}_{|\nu|\mu}$$

We can verify this by direct computation or by noticing that, if we try to use the method of the geodesic coordinate system, we shall come across derivatives of Christoffel symbols which will neither vanish nor cancel out. Only in special spaces will the commutative property hold; this subject will be examined further in Chap. 5.

3.2 Applications of Tensor Analysis

Divergence of a vector. The quantity $\xi^i{}_{\|l}$ is a tensor, so $\xi^i{}_{\|i}$ is a scalar, being the contraction of a tensor; we call it the *divergence* of the vector ξ^i.

$$\xi^i{}_{\|i} = \operatorname{div}\xi = \xi^i{}_{|i} + \begin{Bmatrix} i \\ i\ s \end{Bmatrix}\xi^s$$

We shall show that this expression can be put into a simpler form which allows very easy computation and no longer contains Christoffel symbols.

Indeed, the Ricci theorem proved above gives

$$g_{ij|h} - \begin{Bmatrix} k \\ h\ i \end{Bmatrix} g_{kj} - \begin{Bmatrix} k \\ h\ j \end{Bmatrix} g_{ik} = 0$$

Multiplication and contraction with g^{ij} leads to

$$g^{ij}g_{ij|h} - \begin{Bmatrix} i \\ h\ i \end{Bmatrix} - \begin{Bmatrix} j \\ h\ j \end{Bmatrix} = 0$$

and therefore, since j is a dummy index,

$$\begin{Bmatrix} i \\ i\ h \end{Bmatrix} = \tfrac{1}{2}g^{ij}g_{ij|h}$$

We know that g^{ik} are the elements of the inverse matrix of g_{ik}, and therefore $g^{ik} = \Delta^{ik}/g$, where

$$g = \det((g_{ik}))$$

and Δ^{ik} is the cofactor of g_{ik} as defined in the theory of determinants. But we could expand the determinant g along one of its rows, say the third row, into $g = g_{3k}\,\Delta^{3k}$; from this we see that $\partial g/\partial g_{ik} = \Delta^{ik}$, since Δ^{ik} is a subdeterminant of g which does not contain the variable g_{ik}. Replacing Δ^{ik} by this expression, we have

$$g^{ik} = \frac{1}{g}\frac{\partial g}{\partial g_{ik}}$$

and

$$(3.11) \qquad \begin{Bmatrix} i \\ i\ h \end{Bmatrix} = \frac{1}{2}\frac{1}{g}\frac{\partial g}{\partial g_{ik}}\frac{\partial g_{ik}}{\partial x^h} = \frac{1}{2g}\frac{\partial g}{\partial x^h} = \frac{1}{2}\frac{\partial}{\partial x^h}\log|g|$$
$$= \frac{\partial}{\partial x^h}\log\sqrt{-g}$$

a result which is important in its own right.

This last way of writing a contracted Christoffel symbol of the second kind is legitimate only if the quantity $-g$ under the square-root sign is positive. We shall assume usually that the metric considered has a g_{ik} tensor of signature identical with the one of special relativity $(+---)$. This assumption is usually made in general relativity in the form that one can find, at every point considered, a local coordinate

system in which the g_{ik} tensor reduces to the Lorentz metric tensor of special relativity.

Putting the above results into the divergence formula, we obtain

$$\xi^i{}_{\|i} = \operatorname{div}\xi = \frac{1}{\sqrt{-g}}\left[\xi^i{}_{|i}\sqrt{-g} + \xi^s\frac{\partial\sqrt{-g}}{\partial x^s}\right]$$

that is,

$$(3.12) \qquad \xi^i{}_{\|i} = \operatorname{div}\xi = \frac{1}{\sqrt{-g}}(\xi^i\sqrt{-g})_{|i}$$

We shall see in the next section that the divergence formula is specially suited to formulate Gauss's integral theorem in a Riemann space.

Note that the divergence operation has been defined only on the contravariant form of a vector; this is no restriction, however, since one can always operate with the contravariant components of a given vector.

Geometric interpretation of the divergence formula. Let us consider a change of coordinates from x^i to $\bar{x}^i$. The metric tensor g_{ik} transforms according to $\bar{g}_{ik} = \dfrac{\partial x^\alpha}{\partial\bar{x}^i}\dfrac{\partial x^\beta}{\partial\bar{x}^k}g_{\alpha\beta}$. Therefore, using the rules of matrix multiplication, we see that the determinant g transforms according to

$$\bar{g} = g\left(\det\frac{\partial x^\gamma}{\partial\bar{x}^i}\right)^2$$

and hence, since we always choose the positive value of the square root,

$$\sqrt{-\bar{g}} = \sqrt{-g}\left|\frac{\partial(x^1, x^2, \ldots)}{\partial(\bar{x}^1, \bar{x}^2, \ldots)}\right|$$

where the last factor is the inverse of the absolute value of the Jacobian of the coordinate transformation. On the other hand, one knows that the differential volume element, say, in four dimensions,

$$d\tau = dx^0\,dx^1\,dx^2\,dx^3$$

transforms according to

$$d\bar{\tau} = \left|\frac{\partial(\bar{x}^1, \bar{x}^2, \ldots)}{\partial(x^1, x^2, \ldots)}\right| d\tau$$

Thus

$$\sqrt{-\bar{g}}\, d\bar{\tau} = \sqrt{-g}\, d\tau$$

and therefore $\sqrt{-g}\, d\tau$ is a scalar. We shall call it the invariant four-dimensional volume element. As mentioned before, we assume that we can find always at each point of the Riemann space a coordinate system such that the g_{ik} tensor takes the form of the usual Lorentz metric tensor of special relativity

$$\begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

In such a coordinate system the invariant four-dimensional volume element $\sqrt{-g}\, d\tau$ becomes $d\tau = dx_0\, dv$, the "natural" volume element in that local system of reference in which a physicist measures lengths with rods and time with clocks.

Let us now consider the integral

$$\int_D (\operatorname{div} \xi)(\sqrt{-g}\, d\tau) = \int_D \xi^i{}_{\|i} \sqrt{-g}\, d\tau$$

over an n-dimensional domain D, where we have explicitly displayed the invariant volume element introduced above. This integral is obviously an invariant scalar quantity. With the help of the divergence formula (3.12) we can write it in the form

$$\int_D (\xi^i \sqrt{-g})_{|i}\, d\tau$$

which becomes, by direct application of Green's formula (integration by parts),

$$(3.13)\qquad \int_D (\xi^i \sqrt{-g})_{|i}\, d\tau = \sum_i \int_{\text{boundary}} \xi^i \sqrt{-g}\, dx^1 \cdots dx^{i-1}\, dx^{i+1} \cdots dx^n$$

We interpret expression (3.13) as the flux of the vector ξ^i across the boundary of the domain D. The identity (3.13) is called *Gauss's integral theorem* for a Riemann space. We see that the integral of a divergence of a vector in a volume depends only on the values of the vector on the

boundary, just as in the Euclidean space of ordinary integral calculus. This formula will be very useful later to express conservation laws in the theory of general relativity.

Laplacian of a scalar field. Let us consider the divergence of the gradient of a scalar function $W(x^\alpha)$. We know that this operation leads to the Laplacian in ordinary calculus. The gradient of $W(x^\alpha)$ is a covariant vector $W(x^\alpha)_{|k}$. We form its contravariant components by multiplying by g^{ik} and take its divergence:

$$(3.14)\qquad \frac{1}{\sqrt{-g}} (\sqrt{-g}\, g^{ik} W_{|k})_{|i} = \nabla^2 W = (g^{ik} W_{|k})_{\|i}$$

In analogy with the Euclidean case, we call this scalar the Laplacian of W.

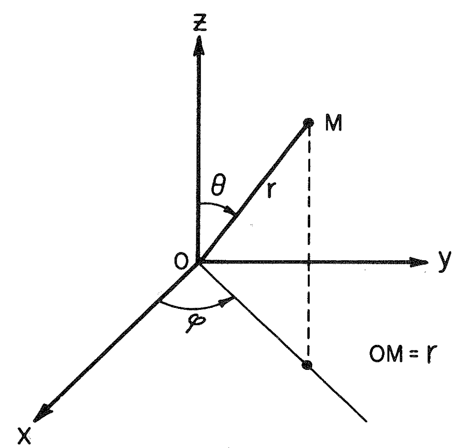


Fig. 3.2

The present formalism can be used to obtain very easily the expression for the Laplacian in any curvilinear coordinate system in ordinary analysis. Let us take as example the case of polar coordinates in *three-dimensional* Euclidean space: r, θ, φ, as indicated in Fig. 3.2. The Euclidean metric in these coordinates is

$$ds^2 = dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2$$

Thus g_{ik} is diagonal, and

$$g_{11} = 1 \qquad g_{22} = r^2 \qquad g_{33} = r^2 \sin^2\theta$$

Therefore $g = r^4 \sin^2\theta$ and

$$g^{11} = 1 \qquad g^{22} = \frac{1}{r^2} \qquad g^{33} = \frac{1}{r^2 \sin^2\theta}$$

Here g is positive, and therefore we simply replace $\sqrt{-g}$ by $\sqrt{g}$ in (3.14) to obtain the expression for the Laplacian in the present case. The Laplacian of a function $W(r,\theta,\varphi)$ is therefore

$$\nabla^2 W = \frac{1}{\sqrt{g}}\left[\sqrt{g}\left(g^{i1}\frac{\partial W}{\partial r} + g^{i2}\frac{\partial W}{\partial \theta} + g^{i3}\frac{\partial W}{\partial \varphi}\right)\right]_{|i}$$

$$\nabla^2 W = \frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}\left(r^2\sin\theta\frac{\partial W}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(r^2\sin\theta\frac{1}{r^2}\frac{\partial W}{\partial\theta}\right)\right.$$
$$\left.+ \frac{\partial}{\partial\varphi}\left(r^2\sin\theta\frac{1}{r^2\sin^2\theta}\frac{\partial W}{\partial\varphi}\right)\right]$$

which becomes, after simplifications,

$$\nabla^2 W = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial W}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial W}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 W}{\partial\varphi^2}$$

We thus obtain this well-known formula in a straightforward way by using the powerful coordinate invariant formalism of tensor analysis.

3.3 Symmetric and Antisymmetric Tensors

We introduce two classes of tensors which play an important role in physical theories. They are defined in the following way:

1. A totally symmetric tensor† is a tensor whose components remain unchanged under interchange of any two indices:

$$T_{ikl} = T_{kil} = T_{lik}$$

2. A totally antisymmetric tensor† is a tensor whose components change sign under any odd permutation of its indices. Clearly, it will therefore remain unchanged under an even permutation of indices:

$$T_{ikl} = -T_{kil} = T_{kli}$$

† Totally symmetric and totally antisymmetric tensors will be simply called symmetric and antisymmetric in further developments.

The simplest example of a symmetric tensor is the metric tensor g_{ik}. The most important antisymmetric tensor in physics is probably the well-known Maxwell tensor, which will be studied in detail in Chap. 4. A justification for singling out the symmetric or antisymmetric character of tensors rests on the following theorem.

Theorem. The symmetric and antisymmetric character of a tensor is an intrinsic property; that is, the symmetry characteristics do not depend upon the coordinate system used.

Proof. Given the tensor $T_{\alpha\beta}$ in one coordinate system x^i, consider another coordinate system $\bar{x}^i$ in which the tensor has components

$$\bar{T}_{ik} = \frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial x^\beta}{\partial \bar{x}^k} T_{\alpha\beta}$$

and

$$\bar{T}_{ki} = \frac{\partial x^\beta}{\partial \bar{x}^k}\frac{\partial x^\alpha}{\partial \bar{x}^i} T_{\beta\alpha}$$

We see that, if $T_{\alpha\beta} = T_{\beta\alpha}$, then $\bar{T}_{ik} = \bar{T}_{ki}$, and if $T_{\alpha\beta} = -T_{\beta\alpha}$, then $\bar{T}_{ik} = -\bar{T}_{ki}$, which proves the theorem for tensors of rank 2. The generalization to tensors of arbitrary rank is immediate.

Notice that, in the three-dimensional space, an antisymmetric tensor of rank 2 has three independent components with which a vector can be associated. Indeed, let ξ^i and η^i be two arbitrary vectors and form the tensor $\xi^i\eta^k - \xi^k\eta^i$. Its three independent components coincide with the components of the exterior or vector product of ξ^i and η^i. Thus, in many applications in three-dimensional space, antisymmetric tensors of this form are identified with vectors. There is, however, one simple way of distinguishing between genuine vectors and such apparent vectors. If all coordinate axes are reversed in direction, each vector ξ^i goes over into $-\xi^i$, while the antisymmetric tensor $\xi^i\eta^k - \xi^k\eta^i$ remains unchanged. This difference in transformation behavior led to the distinction in classical vector theory between polar and axial vectors.

In a four-dimensional space a symmetric tensor of the second rank has 10 independent components, whereas an antisymmetric tensor of the second rank has 6 independent components and is sometimes called a six-vector. For example, the independent components of the Maxwell field tensor are the components of the electric and magnetic fields **E** and **H**. An antisymmetric tensor of the fourth rank has only one independent component; so has any antisymmetric tensor of rank equal to the dimension of the space considered. Antisymmetric tensors of rank higher than the space dimension are identically zero.

Curl of a vector. An antisymmetric tensor field can be obtained from a vector field in the following manner. Consider a covariant vector η_i and form the expressions

$$\eta_{i\|k} = \eta_{i|k} - \left\{ {r \atop i\ k} \right\} \eta_r$$

$$\eta_{k\|i} = \eta_{k|i} - \left\{ {r \atop k\ i} \right\} \eta_r$$

The Christoffel symbols are symmetric in their lower indices. Therefore the difference of the two tensors written above gives the tensor

$$\eta_{i\|k} - \eta_{k\|i} = \eta_{i|k} - \eta_{k|i}$$

which does *not* involve Christoffel symbols. This tensor is obviously antisymmetric; we call it the curl or rotation of the vector η_i. Note that the curl operation can be performed only on the covariant components of a vector, since, in the above, we made use of the symmetry of the Christoffel symbols in its two lower indices.

Homogeneous forms and symmetry character of tensors. The study of antisymmetric tensors is linked directly with the theory of exterior differential forms, which we shall introduce now. From the point of view of integration in an n-dimensional Riemann space, a differential volume element $dx^1\,dx^2 \cdots dx^n$ is not independent of the order of the factors. Indeed, $dx^1\,dx^2 \cdots dx^n$ is not equivalent to $dx^2\,dx^1 \cdots dx^n$, because going from one integration scheme to the other involves the change of coordinates

$$\bar{x}^1 = x^2$$

$$\bar{x}^2 = x^1$$

from which we get, by computing the Jacobian of the transformation,

$$dx^2\,dx^1 = d\bar{x}^1\,d\bar{x}^2 = \frac{\partial(\bar{x}^1,\bar{x}^2)}{\partial(x^1,x^2)}\,dx^1\,dx^2 = -dx^1\,dx^2$$

We see that the differential elements appear anticommutative from the point of view of integration; this simply reflects the change of orientation of the surface element $dx^1 \wedge dx^2$ when one changes the order of integration. We introduce in the space of differential elements the concept of exterior product, which preserves the laws of algebra, except that commutativity is replaced by anticommutativity. The symbol for exterior multiplication is $\wedge$. We thus have

$$dx^i \wedge dx^k = -dx^k \wedge dx^i$$

which in particular leads to

$$dx^i \wedge dx^i = 0$$

This notion goes back to Grassmann (Grassmann, 1878) and is most familiar in the vector product of elementary vector algebra. As we saw above, the exterior product will play a major role in integration theory. However, at this point we shall utilize only the fact that a totally antisymmetric tensor can be best defined as the coefficient system of an exterior differential form. We define an exterior differential form as a form of differentials in which multiplication is understood as exterior multiplication; for example,

$$G = A_{ik}\,dx^i \wedge dx^k$$

is an exterior differential form of order 2. One knows that any usual form $L = B_{ikl}x^ix^kx^l$ can be written with totally symmetric coefficients, since if B_{ikl} is not totally symmetric in i, k, l, we replace it by

$$C_{ikl} = \tfrac{1}{6}(B_{ikl} + B_{kil} + B_{lki} + B_{ilk} + B_{lik} + B_{kli})$$

which gives the same value to L and is symmetric in i, k, l. The study of such forms L reduces, therefore, to the study of forms with symmetric coefficients. The analogous property for exterior differential forms is that one can always utilize coefficients which are antisymmetric in all indices.

Let us show this in the case of two indices. Consider a form

$$G = A_{ik}\,dx^i \wedge dx^k$$

where A_{ik} is arbitrary. Because of our freedom in the choice of dummy indices, we can write it as $G = \frac{1}{2}(A_{ik}\,dx^i \wedge dx^k + A_{ki}\,dx^k \wedge dx^i)$ by simply relabeling summation indices. But the above expression may be written

$$G = \tfrac{1}{2}(A_{ik} - A_{ki})\,dx^i \wedge dx^k$$

by virtue of the anticommutativity of dx^i and dx^k. Since $(A_{ik} - A_{ki})$ is obviously antisymmetric, this proves our original proposition.

By considering an exterior differential form built up from an arbitrary tensor A_{ikl}, we can construct a new tensor which is antisymmetric. Consider, for example, $F = A_{ikl}\, dx^i \wedge dx^k \wedge dx^l$. It can be written

$$F = \frac{1}{3!}(A_{ikl} - A_{kil} + A_{kli} - A_{ilk} + A_{lik} - A_{lki})\, dx^i \wedge dx^k \wedge dx^l$$

That is, we sum over all permutations of the indices, giving a plus sign to an even permutation and a minus sign to an odd permutation. We thus define

$$F = \{A_{ikl}\}\, dx^i \wedge dx^k \wedge dx^l$$

in which the notation $\{A_{ikl}\}$ designates the antisymmetrized sum in the above sense over all permutations of the indices i, k, l:

$$\{A_{ikl}\} = \frac{1}{3!}(A_{ikl} - A_{kil} + A_{kli} - A_{ilk} + A_{lik} - A_{lki}) \tag{3.15}$$

The extension of this definition to an arbitrary number of indices is obvious.

Remarks

1. The antisymmetrization operation { } is linear: one verifies immediately from the above definition that

$$\{A_{ikl} + B_{ikl}\} = \{A_{ikl}\} + \{B_{ikl}\}$$

2. Any tensor of rank 2 can be uniquely decomposed into the sum of a symmetric and an antisymmetric tensor. For tensors of rank higher than 2, this is not possible; in these cases the sum of a symmetric and an antisymmetric tensor does not give enough independent components for the representation of an arbitrary tensor. However, an arbitrary tensor of rank larger than 2 can be decomposed into the sum of more than two tensors whose components have distinguished symmetry properties upon particular permutation of the indices.

3. As a particular case we include tensors of rank 1 as antisymmetric tensors. The antisymmetrization operation is then the identity by definition.

Creation of a new antisymmetric tensor of rank $n + 1$ from an antisymmetric tensor of rank n. In Sec. 3.3 we created an antisymmetric tensor from a vector; we shall here generalize this process. The remarkable feature of this procedure is that the new tensor does not contain Christoffel symbols. Let us first consider the case $q = 3$ for convenience of notation. Given an antisymmetric tensor A_{ikl}, consider its covariant derivative

$$A_{ikl\|m} = A_{ikl|m} - \left\{ {r \atop i\ m} \right\} A_{rkl} - \left\{ {r \atop k\ m} \right\} A_{irl} - \left\{ {r \atop l\ m} \right\} A_{ikr}$$

This tensor is in general not antisymmetric because of the distinguished role of the index m. We antisymmetrize it with the method sketched above, forming

$$A_{ikl\|m}\, dx^i \wedge dx^k \wedge dx^l \wedge dx^m = \{A_{ikl\|m}\}\, dx^i \wedge dx^k \wedge dx^l \wedge dx^m$$

The new antisymmetric tensor $\{A_{ikl\|m}\}$ does not depend upon Christoffel symbols. This results from the following theorem.

Theorem. For every tensor field one has

$$\{A_{ikl\|m}\} = \{A_{ikl|m}\}$$

Proof. By definition, we can write

$$\{A_{ikl\|m}\}\, dx^i \wedge dx^k \wedge dx^l \wedge dx^m = \left[A_{ikl|m} - \left\{ {r \atop i\ m} \right\} A_{rkl} - \left\{ {r \atop k\ m} \right\} A_{irl} - \left\{ {r \atop l\ m} \right\} A_{ikr} \right] dx^i \wedge dx^k \wedge dx^l \wedge dx^m$$

Since each Christoffel symbol is symmetric in its lower indices whereas the differential elements anticommute, this last expression is equal to $A_{ikl|m}\, dx^i \wedge dx^k \wedge dx^l \wedge dx^m$. But, by definition,

$$A_{ikl|m}\, dx^i \wedge dx^k \wedge dx^l \wedge dx^m = \{A_{ikl|m}\}\, dx^i \wedge dx^k \wedge dx^l \wedge dx^m$$

Therefore

$$\{A_{ikl\|m}\} = \{A_{ikl|m}\} \tag{3.16}$$

Note that this process of creation of tensors by differentiation and antisymmetrization always has to start with completely covariant tensors.

By repeating the above procedure, one might think that one can keep creating new antisymmetric tensors of higher ranks. This is not the case, and the above procedure leads to a tensor identically zero at the second step. We illustrate this property in the following examples.

1. Let us begin with a tensor of rank zero, that is, a scalar W. Take its covariant derivative, and do not antisymmetrize since there is only one index present; one obtains the gradient of W, $W_{|k}$. Now repeat the previous operation: Take the covariant derivative of the gradient, antisymmetrize to get the tensor $\{W_{|k|l}\} = \frac{1}{2}(W_{|k|l} - W_{|l|k})$, which is visibly identically zero. We have found the well-known fact that the curl of a gradient vanishes. Thus observe the resultant sequence

$$W; \quad W_{|k}; \quad \{W_{|k|l}\} = 0$$

2. Consider the case of a tensor of rank $q = 1$, that is, the covariant vector ξ_i. Taking its covariant derivative and antisymmetrizing, we obtain the tensor

$$\{\xi_{i\|k}\} = \{\xi_{i|k}\} = \tfrac{1}{2}(\xi_{i|k} - \xi_{k|i})$$

which is proportional to the curl of the vector ξ_i. We repeat the previous operation. Taking the covariant derivative of the curl and antisymmetrizing, we get, in view of the linearity of the antisymmetrization operation,

$$\{\{\xi_{i|k}\}_{|l}\} = \tfrac{1}{2}\{(\xi_{i|k} - \xi_{k|i})_{|l}\} = \tfrac{1}{2}[\{\xi_{i|k|l}\} - \{\xi_{k|i|l}\}]$$

But we note that $\{\xi_{i|k|l}\} = -\{\xi_{k|i|l}\}$, since this antisymmetrized tensor is antisymmetric in i and k. Therefore

$$\{\{\xi_{i|k}\}_{|l}\} = \{\xi_{i|k|l}\}$$

On the other hand, since the order of differentiation is immaterial, we have $\xi_{i|k|l} = \xi_{i|l|k}$, and thus $\{\xi_{i|k|l}\} = \{\xi_{i|l|k}\}$. Since a cyclic permutation leaves the antisymmetrized expression invariant, we obtain also

$$\{\xi_{i|k|l}\} = \{\xi_{k|i|l}\}$$

On the other hand, the antisymmetry in i and k of $\{\xi_{i|k|l}\}$ implies

$$\{\xi_{i|k|l}\} = -\{\xi_{k|i|l}\}$$

Comparing the last two equations, we find

$$\{\xi_{i|k|l}\} = \{\{\xi_{i|k}\}_{|l}\} = 0 \tag{3.17}$$

Again the second antisymmetric differentiation leads to a null tensor.

3. In the case of a second-rank antisymmetric tensor t_{ik}, we consider $t_{ik|l}$ and antisymmetrize. As one can verify immediately, we obtain

$$\{t_{ik|l}\} = \tfrac{1}{3}(t_{ik|l} + t_{kl|i} + t_{li|k}) \tag{3.18}$$

This is an antisymmetric tensor of rank 3 which visibly does not contain any Christoffel symbols. The operation of antisymmetric differentiation on an antisymmetric tensor of rank 2 is analogous to the curl operation on a vector. (In the next chapter we shall see that the Maxwell field tensor satisfies the equations $\{t_{ik|l}\} = 0$, which are equivalent to some of Maxwell's equations involving a curl operation.)

If we now try to repeat the previous operation, we obtain $\{\{t_{ik|l}\}_{|m}\} = 0$. This result is general. In fact, our demonstration for the second example can be extended as it stands to any antisymmetric tensor $A_{jm\cdots npi}$ by forming $\{A_{jm\cdots npi|k}\}$ and proving that

$$\{\{A_{jm\cdots npi|k}\}_{|l}\} = 0 \tag{3.19}$$

since one sees that the indices $jm \cdots np$ are not relevant to the proof. This shows that the number of antisymmetric tensors which can be created by the above process is limited. We can, in fact, create only one such tensor from a given antisymmetric tensor of any rank; repeating the operation gives a null tensor.

3.4 Closed and Exact Tensors

It is a well-known property in ordinary vector analysis that a vector field whose curl is zero is derivable from a potential, and vice versa. We shall introduce here a similar notion into tensor analysis in a Riemann space.

We make the following *definitions:*

1. An antisymmetric tensor t_{ikl} is called a *closed* antisymmetric tensor if

$$\{t_{ikl|n}\} = 0$$

2. An antisymmetric tensor t_{ikl} is called an *exact* antisymmetric tensor if there exists a tensor T_{ik} such that

$$t_{ikl} = \{T_{ik|l}\}$$

T_{ik} is called the tensor potential of t_{ikl}.

Analogously, closed and exact tensors can be defined for every rank. In view of the foregoing definition, the result of the previous section can be formulated as a theorem.

Theorem. *Every exact tensor is closed.*

Indeed, take an antisymmetric tensor T_{ik} as potential and consider $\{T_{ik|l}\}$, which is by definition exact; we know that $\{\{T_{ik|l}\}_{|m}\} = 0$, which proves that $\{T_{ik|l}\}$ is closed. We shall show in the following theorem that the converse proposition is also true.

Theorem. *Every closed tensor is exact,* that is, admits a tensor potential.

We first illustrate this theorem with the familiar case of a vector in Euclidean geometry. If a vector t_i has zero curl, $t_{i|k} - t_{k|i} = 0$, we know that it is derivable from a potential ϕ:

$$t_i = \frac{\partial \phi}{\partial x^i} \tag{3.20}$$

The curl condition represents simply the integrability conditions of (3.20) and allows the quantity

$$d\phi = \frac{\partial \phi}{\partial x^i}\, dx^i = t_i\, dx^i$$

to be an exact differential.

We now prove the above theorem for tensors of rank 2 in an n-dimensional space. The proof will be by induction on the dimension n. We start with the case $n = 2$ and consider an antisymmetric tensor

$$t_{ik} = \begin{pmatrix} 0 & t_{12} \\ -t_{12} & 0 \end{pmatrix}$$

Let us form $\{t_{ik|l}\}$; it is identically zero because i, k, l can take only the values 1 and 2, and therefore there is always one index, 1 or 2, which is present twice. It follows that the result of the antisymmetrization is zero. This is an example of an antisymmetric tensor of rank higher than the space dimension and which must therefore be identically zero, as mentioned earlier. In two dimensions a second-rank antisymmetric tensor is always closed. Let us show that such a closed tensor is exact. We must find out if two functions t_1 and t_2 exist such that

$$t_{12} = \frac{\partial t_1}{\partial x^2} - \frac{\partial t_2}{\partial x^1}$$

This is obviously so, for one can take $t_2 = 0$ and $t_1 = \int t_{12}\, dx^2$. Therefore every second-rank antisymmetric tensor in two dimensions is exact. Before extending this proof to higher dimensions, let us first make the following remark: If we suppose that t_{12} depends on a parameter p and is k times differentiable in p, then, from the known laws of differentiation under the integral sign, the tensor potential which we defined above depends on p with the same differentiability properties.

To complete the induction, we go from dimension n to $n + 1$: Suppose that the indices i, k take the values $1, 2, \ldots, n + 1$ and that t_{ik} is a closed tensor in an $(n + 1)$-dimensional space:

$$\{t_{ik|l}\} = 0 \qquad \text{for} \qquad i, k, l = 1, 2, \ldots, n + 1$$

By our induction hypothesis we can now assert that the matrix t_{ik} considered as a function of x^1 to x^n, and for every fixed x^{n+1}, has a potential t_i such that

$$t_{ik} = t_{i|k} - t_{k|i} \qquad \text{for} \qquad i, k = 1, 2, \ldots, n$$

By the induction hypothesis again, these t_i depend on the parameter x^{n+1} in a differentiable manner. Therefore there exist functions $t_i(x^1, x^2, \ldots, x^n; x^{n+1})$ such that

$$t_{ik} = t_{i|k} - t_{k|i} \qquad \text{for} \qquad i, k = 1, 2, \ldots, n \tag{3.21}$$

To complete the induction we need only prove that we can write

$$t_{ik} = t_{i|k} - t_{k|i}$$

for i or k equal to $n + 1$.

The closure condition on t_{ik} for $l = n + 1$ and $i, k = 1, \ldots, n$ is $\{t_{ik|n+1}\} = 0$. It can be written out as

$$\frac{\partial t_{ik}}{\partial x^{n+1}} + \frac{\partial t_{k,n+1}}{\partial x^i} + \frac{\partial t_{n+1,i}}{\partial x^k} = 0$$

Using the potential representation (3.21) of t_{ik} for $i, k \leq n$, we obtain

$$\frac{\partial t_{i|k}}{\partial x^{n+1}} - \frac{\partial t_{k|i}}{\partial x^{n+1}} + \frac{\partial t_{k,n+1}}{\partial x^i} + \frac{\partial t_{n+1,i}}{\partial x^k} = 0$$

Taking into account the definition of $t_{i|k} = \partial t_i/\partial x^k$ and the fact that t_{ik} is antisymmetric, we arrive at

$$\frac{\partial t_{i|n+1}}{\partial x^k} - \frac{\partial t_{k|n+1}}{\partial x^i} - \frac{\partial t_{n+1,k}}{\partial x^i} + \frac{\partial t_{n+1,i}}{\partial x^k} = 0$$

which can be written

$$\frac{\partial}{\partial x^k}(t_{i|n+1} + t_{n+1,i}) = \frac{\partial}{\partial x^i}(t_{k|n+1} + t_{n+1,k})$$

If we define $W_i = t_{i|n+1} + t_{n+1,i}$, we have $(\partial/\partial x^k)W_i - (\partial/\partial x^i)W_k = 0$, which means that the curl of W_i is 0. In the present case of a vector we know that this condition implies that W_i admits a potential which we call t_{n+1}. We therefore write

$$W_i = \frac{\partial t_{n+1}}{\partial x^i} = t_{n+1|i} \qquad i = 1, 2, \ldots, n$$

which with the definition of W_i gives

$$t_{n+1,i} = t_{n+1|i} - t_{i|n+1}$$

This is the expression we were looking for, analogous to (3.21), but for $k = n + 1$.

We have used a fixed coordinate system x^i and found in it $n + 1$ functions $t_k(x_1, \ldots, x_{n+1})$ such that

$$t_{ik} = t_{i|k} - t_{k|i} \qquad i = 1, \ldots, n + 1$$

We allow the t_k to transform as covariant vectors under a change of coordinates, so that this equation becomes a tensor identity and is therefore valid independently of the coordinate system.

We have thus completed the proof that every closed tensor of rank 2 is exact. This property and its converse are true for tensors of arbitrary rank, but we shall make use of it only in the case of a tensor of rank 2 in four-dimensional space when writing Maxwell's equations in tensor form in the next chapter.

3.5 Tensor Densities—Dual Tensors

When studying the properties of the divergence of a vector under integration, we introduced the expression $\sqrt{-g}$ in order to be able to form

invariant integrals. This expression is a special case of an entity which plays an important role in general relativity, the so-called tensor density. A tensor density (which is usually denoted by a script letter such as $\mathfrak{T}_{\alpha\beta}{}^{\gamma}$) depends on the coordinate system in such a way that, under change of variables $x^\alpha \to \bar{x}^\alpha$, it obeys the transformation law

$$\bar{\mathfrak{T}}_{\alpha\beta}{}^{\gamma} = \frac{\partial x^k}{\partial \bar{x}^\alpha}\frac{\partial x^l}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^\gamma}{\partial x^m}\mathfrak{T}_{kl}{}^{m}\frac{\partial(x^1, x^2, \ldots)}{\partial(\bar{x}^1, \bar{x}^2, \ldots)} \tag{3.22}$$

In order to avoid sign difficulties we shall restrict ourselves to transformations with a positive Jacobian. Then, in the sense of this definition, $\sqrt{-g}$ may be called a scalar density. This name comes from the fact that, if one integrates such a quantity, forming, for instance, the integral $I = \int \sqrt{-g}\, d\tau$, the result is an invariant as we saw earlier; thus $\sqrt{-g}$ behaves like the physical density in space of the quantity I.

If $T_{\alpha\beta}{}^{\gamma}$ is a tensor, clearly

$$\mathfrak{T}_{\alpha\beta}{}^{\gamma} = T_{\alpha\beta}{}^{\gamma}\sqrt{-g} \tag{3.23}$$

is a tensor density. The transition from tensor to tensor density can always be performed by a correspondence (3.23), and hence the knowledge of $\sqrt{-g}$ leads to a complete understanding of all tensor densities.

As illustration, let us consider in four-dimensional space four arbitrary contravariant vectors $\xi^\alpha_{(j)}$, $j = 1, 2, 3, 4$, and form the determinant

$$D = \det(\xi^\alpha_{(j)}) = \epsilon_{\alpha\beta\gamma\delta}\xi^\alpha_{(1)}\xi^\beta_{(2)}\xi^\gamma_{(3)}\xi^\delta_{(4)} \tag{3.24}$$

D appears here as a multilinear form of the four vectors with the coefficient system

$$\epsilon_{\alpha\beta\gamma\delta} = \begin{cases} 0 & \text{if any two indices are equal} \\ \pm 1 & \text{according to } (\alpha\beta\gamma\delta)\text{, being an even or odd permutation of the numbers } (1,2,3,4) \end{cases}$$

The ϵ-system is well known from elementary determinant theory. To keep the symmetry between upper and lower indices in tensor calculus, we also introduce the symbol $\epsilon^{\alpha\beta\gamma\delta}$, which is exactly the same numerical array as $\epsilon_{\alpha\beta\gamma\delta}$ but allows us to keep Einstein's summation convention when dealing with covariant vector components.

In our geometry the ϵ-system provides a multilinear form of vectors with a very simple transformation behavior under change of coordinates.

Indeed, from

$$\bar{\xi}^{\alpha}_{(j)} = \frac{\partial \bar{x}^{\alpha}}{\partial x^{l}} \xi^{l}_{(j)}$$

and the multiplication rule for determinants, we find that

$$\bar{D} = \frac{\partial(\bar{x}^1, \bar{x}^2, \ldots)}{\partial(x^1, x^2, \ldots)} D$$

Thus D^{-1} is a scalar density and $D\sqrt{-g}$ is a proper scalar. Multiplying both sides of (3.24) by $\sqrt{-g}$ and applying the quotient theorem to $D\sqrt{-g}$ and the four arbitrary vectors $\xi_{(j)}$, we find that

$$\epsilon_{\alpha\beta\gamma\delta}\sqrt{-g} = e_{\alpha\beta\gamma\delta} \tag{3.25a}$$

is an antisymmetric covariant tensor of rank 4, the Levi-Civita tensor. As remarked earlier, in Sec. 3.3, in a four-dimensional space, **e** is the only such antisymmetric tensor (within a multiplicative factor). The covariant components of **e** can be obtained from similar reasoning applied to covariant vector components; they are, as an easy calculation shows,

$$e^{\alpha\beta\gamma\delta} = \frac{-1}{\sqrt{-g}} \epsilon^{\alpha\beta\gamma\delta} \tag{3.25b}$$

We see that $\epsilon^{\alpha\beta\gamma\delta}$ is a tensor density in the sense of (3.23).

While the tensor density $\epsilon^{\alpha\beta\gamma\delta}$ has components which are independent of the coordinate system (similar to the Kronecker tensor $g^{i}{}_{k} = \delta^{i}{}_{k}$), it should be observed that the tensors $e^{\alpha\beta\gamma\delta}$ and $e_{\alpha\beta\gamma\delta}$ both have zero covariant derivatives. Indeed,

$$e_{\alpha\beta\gamma\delta\|l} = \epsilon_{\alpha\beta\gamma\delta}(\sqrt{-g})_{|l} - \begin{Bmatrix} r \\ \alpha\ l \end{Bmatrix} e_{r\beta\gamma\delta} - \begin{Bmatrix} r \\ \beta\ l \end{Bmatrix} e_{\alpha r\gamma\delta} - \begin{Bmatrix} r \\ \gamma\ l \end{Bmatrix} e_{\alpha\beta r\delta} - \begin{Bmatrix} r \\ \delta\ l \end{Bmatrix} e_{\alpha\beta\gamma r}$$

However, the four indices in $e_{\alpha\beta\gamma\delta}$ must be different to give nonzero components. Hence, in each Christoffel symbol, r must be identical with the index which it replaces in $e_{\alpha\beta\gamma\delta}$. We find, therefore,

$$e_{\alpha\beta\gamma\delta\|l} = \epsilon_{\alpha\beta\gamma\delta}\left[(\sqrt{-g})_{|l} - \begin{Bmatrix} r \\ r\ l \end{Bmatrix}\sqrt{-g}\right]$$

which is identically zero by virtue of (3.11).

By means of the coordinate independent coefficient system $\epsilon_{\alpha\beta\gamma\delta}$, we can establish a very simple correspondence between antisymmetric tensor densities of rank 2 and antisymmetric tensors of rank 2. Let $\mathfrak{T}^{\alpha\beta}$ be an antisymmetric tensor density,

$$\mathfrak{T}^{\alpha\beta} = T^{\alpha\beta}\sqrt{-g}$$

We define

$$(^*T)_{\alpha\beta} = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}\mathfrak{T}^{\gamma\delta} = \tfrac{1}{2}e_{\alpha\beta\gamma\delta}T^{\gamma\delta} \tag{3.26}$$

which is an antisymmetric covariant tensor of rank 2. We call $^*T_{\alpha\beta}$ the *dual tensor* of the tensor $T^{\gamma\beta}$. Clearly, the tensor component $^*T_{\alpha\beta}$ coincides with the tensor density component $\mathfrak{T}^{\gamma\delta}$, with complementary indices, where $(\alpha,\beta,\gamma,\delta)$ form an even permutation of (1,2,3,4). For example,

$$^*T_{12} = \mathfrak{T}^{34} \qquad ^*T_{14} = \mathfrak{T}^{23} \qquad ^*T_{13} = \mathfrak{T}^{42} = -\mathfrak{T}^{24} \tag{3.27}$$

This fact shows the validity of the inverse formula

$$\mathfrak{T}^{\alpha\beta} = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}(^*T)_{\gamma\delta} \qquad T^{\alpha\beta} = -\tfrac{1}{2}e^{\alpha\beta\gamma\delta}(^*T)_{\gamma\delta} \tag{3.26'}$$

which leads back from the dual tensor to the original one.

The notions of tensor density and dual tensor will be used in the next chapter to write Maxwell's equations in a very condensed form in a four-dimensional space. Furthermore, we shall make use of a property of second-rank antisymmetric tensors, which we now prove.

Theorem. *The two following properties of a second-rank antisymmetric tensor in four-dimensional space are equivalent: being closed and having a zero-divergence dual tensor;* that is,

$$\{T_{\alpha\beta|\lambda}\} = \{T_{\alpha\beta\|\lambda}\} = 0 \quad \text{implies} \quad (^*T^{\mu\nu})_{\|\nu} = 0 \quad \text{and vice versa.}$$

To prove this theorem note that from the definition (3.26) and the fact that the tensor $e_{\alpha\beta\gamma\delta}$ has a zero covariant derivative, we may write

$$(^*T^{\mu\nu})_{\|\nu} = \tfrac{1}{2}e^{\mu\lambda\alpha\beta}T_{\alpha\beta\|\lambda} \tag{3.28}$$

The summation indices λ, α, β, are dummy indices, and we can permute them arbitrarily in the above identity. If we write out all permutations and add the resulting six equations, the symmetry properties of $e^{\mu\lambda\alpha\beta}$ allow us to write the result as

$$(^*T^{\mu\nu})_{\|\nu} = e^{\mu\lambda\alpha\beta}\{T_{\alpha\beta|\lambda}\} \tag{3.29}$$

where the covariant derivative has been replaced by the ordinary derivative as allowed by (3.16). The two expressions in (3.29) are zero or nonzero together, which proves the theorem. Our reasoning also shows that the covariant divergence of an antisymmetric tensor can be expressed without the use of Christoffel symbols, as is evident from (3.29).

Finally, it should be remembered that to form invariant scalar quantities by integration, one must integrate over scalar densities. The expression

$$\int \varphi \sqrt{-g}\, d\tau \tag{3.30}$$

is a scalar if φ is a scalar function. On the other hand, the integral of a tensor density,

$$\int \mathfrak{T}^{\alpha\beta}\, d\tau = \int T^{\alpha\beta}\sqrt{-g}\, d\tau \tag{3.31}$$

has no well-defined transformation properties since it is not attached to a single point in space, and therefore no transformation coefficients $\partial\bar{x}^\alpha/\partial x^\beta$ can be defined. Only in the limiting case, when the volume τ shrinks down to an *infinitesimal neighborhood* of a given point P, can we say that (3.31) has meaning, for then the transformation coefficients $\partial\bar{x}^\alpha/\partial x^\beta$ are definable at P, and

$$\int \bar{\mathfrak{T}}^{\alpha\beta}\, d\bar{\tau} = \left(\frac{\partial\bar{x}^\alpha}{\partial x^\gamma}\right)_P \left(\frac{\partial\bar{x}^\beta}{\partial x^\delta}\right)_P \int \mathfrak{T}^{\gamma\delta}\, d\tau$$

is the transformation law for the integral.

3.6 Vector Fields on Curves

We introduced in Sec. 3.1 the concept of covariant derivative $\xi^i_{\|k}$ for a vector field $\xi^i(x^j)$ and used it to create from a given vector field new tensor fields by covariant differentiation. The basic idea in this operation is the comparison of the local change of the component ξ^i due to the form of the function $\xi^i(x^j)$ with the corresponding change according to the law of

vector transplantation. The difference of these changes measures the absolute variation of the vector field and gives rise to covariant expressions.

It is natural to apply the same method in the case of a vector field which is defined only on a curve. Let $x^i(s)$ be the parametric representation of a curve Λ, and consider a vector field $\xi^i(s)$ given as a function of the curve parameter. If we move from $x^i(s)$ to $x^i(s + \Delta s)$, the vector components will change by

$$\xi^i(s + \Delta s) = \xi^i(s) + \frac{d\xi^i}{ds}\Delta s + 0(\Delta s^2) \tag{3.32}$$

while the vector transplantation (3.2) of $\xi^i(s)$ along the curve would have led to the vector

$$\xi^{i*}(s + \Delta s) = \xi^i(s) + \Gamma^i_{kl}\xi^l(s)\frac{dx^k}{ds}\Delta s + 0(\Delta s^2) \tag{3.33}$$

Hence we define the absolute derivative of the vector field $\xi^i(s)$ along the curve Λ by the formula

$$\frac{D\xi^i(s)}{Ds} = \frac{d\xi^i}{ds} - \Gamma^i_{kl}\frac{dx^k}{ds}\xi^l \tag{3.34}$$

The generalization of this operation to tensor fields given only along a curve is obvious. If $\xi^i(s)$ happens to be the restriction to the curve $x^j(s)$ of a general vector field $\xi^i(x^j)$, we have

$$\frac{D\xi^i}{Ds} = \left(\frac{\partial\xi^i}{\partial x^k} - \Gamma^i_{kl}\xi^l\right)\frac{dx^k}{ds} = \xi^i_{\|k}\frac{dx^k}{ds} \tag{3.35}$$

and more generally for corresponding tensor fields,

$$\frac{D}{Ds}T^{ik}(x^l) = T^{ik}_{\|l}\frac{dx^l}{ds} \tag{3.35$'$}$$

The operation of absolute differentiation creates from vector fields along Λ new vector fields:

$$\xi^i = \frac{D\xi^i}{Ds} \tag{3.36}$$

That this is indeed a vector can be verified from the transformation law

(2.5) for connections and from the definition (3.34). It is now easily seen that

$$\dot{\xi}_i = g_{ik}\dot{\xi}^k = \frac{D}{Ds}(g_{ik}\xi^k) = \frac{D}{Ds}\xi_i \tag{3.36'}$$

since the tensor g_{ik} is constant under absolute differentiation. We also verify easily the laws of product differentiation; for example,

$$\frac{D}{Ds}(U_iV^j) = \dot{U}_iV^j + U_i\dot{V}^j \tag{3.37}$$

The concept of absolute differentiation allows us to develop a differential geometry for curves in a Riemann space, which is quite analogous to the corresponding curve theory in Euclidean space. The most important vector field along a curve is its tangent vector field $t^i(s)$; it is by definition proportional to dx^i/ds. To be more specific, let us suppose that Λ is a timelike curve and that s is its arc-length parameter. We may then take t^i to be a unit vector

$$t^i = \frac{dx^i}{ds} \qquad t^it_i = 1 \tag{3.38}$$

Next we let

$$m^i = \dot{t}^i = \frac{Dt^i}{Ds} \tag{3.39}$$

From the second equation (3.38) we conclude by absolute differentiation

$$m^it_i = m_it^i = 0 \tag{3.40}$$

The vector m^i is orthogonal to the timelike tangent vector t^i and is therefore spacelike. We define a unit vector in the direction of m^i by

$$n^i = \frac{1}{\kappa}m^i \qquad n_in^i = -1$$

and can then write (3.39) in the form

$$\dot{t}^i = \kappa n^i \tag{3.41}$$

The unit vector n^i is the principal normal of Λ, and formula (3.41) is the generalization of Frenet's formula in classical curve theory. It relates the derivative of the tangent vector to the normal vector n^i by means of the principal curvature κ.

We shall not pursue the theory of curves any further, but shall use the two important vector fields $t^i(s)$ and $n^i(s)$ along a curve to define a law of vector transport along a curve which is very similar to the law of vector transplantation treated before. We define the tensor field along Λ by use of the quantities t^i, n^i, and κ defined in (3.41):

$$T^{ik}(s) = \kappa(s)\left[n^i(s)t^k(s) - t^i(s)n^k(s)\right] \tag{3.42}$$

and consider the differential equation defining the vector field $V^i(s)$,

$$\frac{DV^i}{Ds} = T^i{}_k(s)V^k \tag{3.43}$$

By use of (3.34) we can bring (3.43) into the form

$$\frac{dV^i}{ds} = \left[\Gamma^i_{kl}\frac{dx^k}{ds} + T^i{}_l(s)\right]V^l(s) \tag{3.44}$$

which shows that (3.43) is a first-order linear homogeneous differential equation for the unknown vector $V^i(s)$. If we prescribe V^i at one point of Λ, say, for $s = 0$, we can determine the vector field $V^i(s)$ along Λ uniquely by means of (3.44). We may thus conceive (3.43) as a law of vector transport along the curve Λ.

Observe now that the vector $t^i(s)$ satisfies the differential equation (3.43) identically. By substitution, using (3.42) and (3.38), (3.40), and (3.41), we obtain

$$\frac{Dt^i}{Ds} = \kappa\left[n^it_kt^k - t^in_kt^k\right] = \kappa n^i = \dot{t}^i \tag{3.45}$$

which verifies our assertion; that is, the tangent vector $t^i(s)$ is carried along Λ by the special transport law (3.43).

Next, let $V^i(s)$ and $W^i(s)$ be two vector fields on Λ which are transported by the same law (3.43). We easily find

$$\frac{D(V^iW_i)}{Ds} = T^i{}_kV^kW_i + T^i{}_kW^kV_i = T^{ik}(V_iW_k + W_iV_k) = 0 \tag{3.46}$$

because of the antisymmetry of the tensor $T^{ik}(s)$. Thus the transport law (3.43) preserves the scalar product, and hence the length and angles of all vectors so displaced. It is thus closely analogous to the displacement law for general vector fields discussed above.

However, in general, our new transport law (3.43) will be different from the law of parallel displacement valid for the entire Riemann space.

Indeed, the only curves whose tangent vectors are obtained by parallel displacement are the geodesics, while our new transport formula is so constructed that it carries the tangent vector of Λ into itself. On the other hand, the transport law (3.43) is much more specialized than the parallel displacement law and depends strongly on Λ. However, in many physical applications the transport formula (3.43) is of significant value. It may happen that a particular curve in space-time plays a distinguished role without being a geodesic; for example, the timelike world-line of an observer will usually not be a geodesic. It will then be a great convenience to introduce a coordinate system which moves with the observer, preserves all geometrical relations, and has the world-line of the observer as one coordinate axis. The observer will refer his observations to an orthogonal triad of axes in his laboratory which are all three orthogonal in four-space to his four-velocity, thus forming an orthogonal tetrad of reference; he can transport along with him this orthogonal tetrad using the transport law (3.43). It is called the Fermi-Walker transport law. For details and applications, we refer the reader to the bibliography (Fermi, 1922; Pirani, 1957; Synge, 1960; and Walker, 1932).

3.7 Intrinsic Symmetries and Killing Vectors

In general relativity we try to rid ourselves of accidental properties of coordinate frames and are led naturally to the study of tensors. However, many problems possess an intrinsic symmetry which may be difficult to recognize in an arbitrary coordinate system. For example, the problem of describing the gravitational field of a stationary spherical body, which we shall discuss in Chap. 6, possesses intrinsic spherical symmetry; this is evident when the problem is expressed in spherical coordinates but may become hidden if other coordinates are used. We wish to consider how such hidden symmetries can be discovered in an invariant manner when working in an arbitrary coordinate system.

We consider a metric tensor $g_{\alpha\beta}(x^\gamma)$ which admits a one-parameter group of continuous transformations

$$\bar{x}^\mu = \varphi^\mu(x^\gamma;\lambda) \tag{3.47}$$

where λ is the parameter of the group. Under all these changes of markers we assume the metric to be the same. In this sense we can say that the metric has a hidden symmetry. The transformation (3.47) may, for example, represent a rotation group (around a fixed axis since we restricted ourselves to a one-parameter group) in properly chosen coordinates; but in an arbitrary marker system x^γ its appearance must be as general as written down. Under any change of markers we have the transformation formula for the $g_{\mu\nu}$ tensor:

$$\bar{g}_{\mu\nu}(\bar{x}^\kappa)\,\frac{\partial \bar{x}^\mu}{\partial x^\alpha}\frac{\partial \bar{x}^\nu}{\partial x^\beta} = g_{\alpha\beta}(x^\gamma) \tag{3.48}$$

But since we demanded that the metric tensor remain unchanged under the transformations (3.47), we have the symmetry requirement

$$\bar{g}_{\mu\nu}(\bar{x}^\kappa) = g_{\mu\nu}(\bar{x}^\kappa) \tag{3.49}$$

That is, the functional dependence of the tensor $g_{\mu\nu}$ on the markers must be the same before and after the transformation. Thus (3.48) yields

$$g_{\mu\nu}(\varphi^\kappa(x^\gamma;\lambda))\varphi^\mu{}_{|\alpha}\varphi^\nu{}_{|\beta} = g_{\alpha\beta}(x^\gamma) \tag{3.50}$$

We differentiate this identity with respect to the parameter λ. This presupposes, of course, that the continuous group of transformations is differentiable, and we make this assumption. We denote differentiation with respect to λ by a dot over the function considered and define the vector field

$$\Psi^\mu(x^\gamma) = \dot{\varphi}^\mu(x^\gamma;0) \tag{3.51}$$

This is indeed a vector field, since it describes the infinitesimal shift of the point x^γ under an infinitesimal increase of the parameter λ and has thus an intrinsic geometric meaning. The differentiation of (3.50) with respect to λ yields, for $\lambda = 0$,

$$g_{\mu\nu|\kappa}(\varphi^\rho)\Psi^\kappa\varphi^\mu{}_{|\alpha}\varphi^\nu{}_{|\beta} + g_{\mu\nu}(\varphi^\rho)\Psi^\mu{}_{|\alpha}\varphi^\nu{}_{|\beta} + g_{\mu\nu}(\varphi^\rho)\varphi^\mu{}_{|\alpha}\Psi^\nu{}_{|\beta} = 0 \tag{3.52}$$

If we assume that the parameter λ is chosen such that $\lambda = 0$ corresponds to the identity transformation

$$x^\mu = \varphi^\mu(x^\gamma;0) \tag{3.53}$$

we derive by differentiation

$$\delta^\mu{}_\gamma = \varphi^\mu{}_{|\gamma} \qquad \text{for } \lambda = 0 \tag{3.54}$$

Hence (3.52) reduces to the identity

(3.55) $$g_{\alpha\beta|\kappa}\Psi^\kappa + g_{\mu\beta}\Psi^\mu{}_{|\alpha} + g_{\alpha\nu}\Psi^\nu{}_{|\beta} = 0$$

The existence of a hidden symmetry of the metric tensor thus leads us to postulate the existence of a vector field Ψ^μ which satisfies the differential system (3.55). The integrability condition for this system is a differential relation for the metric tensor which is a covariant formulation of the symmetry. Indeed, from the theory of continuous groups it follows that a field of infinitesimal generators Ψ^μ guarantees the existence of an integral group $\varphi^\mu(x^\gamma;\lambda)$ as desired.

We do not enter into the mathematical theory of integration of the system (3.55), but wish only to bring this system into a very elegant and suggestive form. We introduce the covariant vector field

(3.56) $$\Psi_\sigma = g_{\sigma\mu}\Psi^\mu$$

and calculate its covariant derivative:

(3.57) $$\Psi_{\sigma\|\rho} = \Psi_{\sigma|\rho} - \begin{Bmatrix}\mu \\ \sigma\ \rho\end{Bmatrix}\Psi_\mu = \Psi_{\sigma|\rho} - \tfrac{1}{2}(g_{\sigma\mu|\rho} + g_{\rho\mu|\sigma} - g_{\sigma\rho|\mu})\Psi^\mu$$
$$= \tfrac{1}{2}(g_{\sigma\mu|\rho} - g_{\rho\mu|\sigma} + g_{\sigma\rho|\mu})\Psi^\mu + g_{\sigma\mu}\Psi^\mu{}_{|\rho}$$

From this equation we obtain

(3.58) $$\Psi_{\sigma\|\rho} + \Psi_{\rho\|\sigma} = g_{\sigma\rho|\mu}\Psi^\mu + g_{\sigma\mu}\Psi^\mu{}_{|\rho} + g_{\rho\mu}\Psi^\mu{}_{|\sigma}$$

Thus we can express the condition (3.55) in the form

(3.59) $$\Psi_{\sigma\|\rho} + \Psi_{\rho\|\sigma} = 0$$

A vector field which satisfies this equation is called a *Killing vector*, after its discoverer. Using the definition (3.59), we have proved that a necessary condition for *the metric $g_{\mu\nu}$ to have a hidden symmetry is that it admits a Killing vector field $\Psi_\sigma(x^\gamma)$.*

In this development, we have restricted ourselves to a one-parameter group of transformations for the sake of clarity in the exposition. The theory can be easily extended to an n-parameter group of transformations. By a straightforward generalization one verifies that the necessary condition for a metric to admit an n-parameter group of transformations is that it admits n Killing vector fields. Such a group of transformations is often called a *group of motions* of the space with the metric $g_{\mu\nu}$.

One of the most interesting physical symmetries that we shall encounter is time-independence. From the preceding discussion we see that an

invariant characterization of this property is that there must exist a Killing vector field that is timelike. This characterization has meaning even in a coordinate system where x^0 is not a convenient time label and where the metric may depend on x^0. For such a geometry a coordinate system can be found in which the metric is time-independent; it is called *stationary*. (In Chap. 6 we shall discuss also a special case of stationary metric that we shall term *static*. These two terms should not be confused.)

Exercises

3.1 Show that if a vector is parallel-displaced along a geodesic in a Riemann space, its angle with the tangent vector to the geodesic remains unchanged. Assume that the metric is positive-definite; see Exercise 1.5.

3.2 Show that lowering an index of $\xi^i{}_{\|k}$ leads to the expression (3.7) for the covariant derivative of a covariant vector field.

3.3 Let Γ^i_{kl} be a set of symmetric affine connections and *demand* that the metric tensor have a zero covariant derivative. Show that this implies $\Gamma^i_{kl} = -\begin{Bmatrix}i \\ k\ l\end{Bmatrix}$. This is an alternative motivation for working in a Riemann space.

3.4 Consider a vector field w_i on a two-dimensional plane. Show that

$$\int_A \{w_{i|k}\}\, dx^i \wedge dx^k = \oint_C w_j\, dx^j$$

for a closed curve C enclosing the area A. From this it follows that if $\{w_{i|k}\} = 0$, then the line integral is zero. Show that it then follows that the vector w_i has a vector potential ϕ, that is, $w_i = \phi_{|i}$. (This proves the theorem in Sec. 3.4 for rank 1 and two dimensions.)

3.5 Show that the following tensor identity holds by working in a tangent Lorentz space:

$$e_{\alpha\beta\gamma\delta}e^\alpha{}_{\sigma\omega\tau} = -[g_{\beta\sigma}g_{\gamma\omega}g_{\delta\tau} - g_{\beta\omega}g_{\gamma\sigma}g_{\delta\tau} + g_{\beta\omega}g_{\delta\sigma}g_{\gamma\tau} - g_{\beta\sigma}g_{\delta\omega}g_{\gamma\tau}$$
$$+ g_{\gamma\sigma}g_{\delta\omega}g_{\beta\tau} - g_{\delta\sigma}g_{\gamma\omega}g_{\beta\tau}]$$

From this show by contraction that

$$e_{\alpha\beta\gamma\delta}e^{\alpha\beta}{}_{\omega\tau} = -2[g_{\gamma\omega}g_{\delta\tau} - g_{\delta\omega}g_{\gamma\tau}]$$

From this it is easy to show that the dual operation operating twice produces the negative of the original tensor:

$$*(*T^{\alpha\beta}) = -T^{\alpha\beta}$$

3.6 Consider a diagonal metric in four dimensions

$$ds^2 = g_{00}(dx^0)^2 + g_{11}(dx^1)^2 + g_{22}(dx^2)^2 + g_{33}(dx^3)^2$$

Show that the invariant four-volume element is

$$(\sqrt{|g_{00}|}\,dx^0)(\sqrt{|g_{11}|}\,dx^1)(\sqrt{|g_{22}|}\,dx^2)(\sqrt{|g_{33}|}\,dx^3)$$

This illustrates the identification of $\sqrt{|g_{00}|}\,dx^0$ with a physical time interval, $\sqrt{|g_{11}|}\,dx^1$ with the physical space interval in the 1 direction, etc. Write out these intervals explicitly in spherical coordinates and verify that this identification agrees with the geometric picture.

3.7 We call a second-rank tensor traceless if

$$g^{\mu\nu}T_{\mu\nu} = T^{\alpha}_{\alpha} = 0$$

Given an arbitrary second rank tensor $S_{\mu\nu}$ in four dimensions, show that

$$S_{\mu\nu} - \tfrac{1}{4}(S^{\alpha}_{\alpha})g_{\mu\nu}$$

is traceless. From this show that an arbitrary second-rank tensor may be written as the sum of an antisymmetric tensor, a symmetric traceless tensor, and a multiple of the metric tensor.

3.8 Consider the simplest case of a field of Killing vectors. In the space of special relativity the metric is independent of position, so that a translation by a constant four-vector ξ^μ is a symmetry. Write this translation as $x'^\mu = x^\mu + \lambda\xi^\mu$. What is the Killing vector corresponding to this symmetry?

3.9 The nature of the translational Killing vector is quite obvious in the above exercise. Now make a transformation from the Cartesian coordinates to cylindrical coordinates ρ, θ, z where it is not so obvious. What is the translational Killing vector in these coordinates? Verify explicitly that it satisfies the fundamental equation (3.59).

3.10 Consider a metric that is invariant under the translation in time, $x'^\mu = x^\mu + \lambda\delta^\mu{}_0$. Such a metric is clearly independent of x^0. What is the Killing vector? Verify Eq. (3.59) explicitly.

Problems

3.1 An alternative way to introduce the concept of the exterior multiplication symbol $\wedge$ of Sec. 3.3 is to consider a two-dimensional surface labeled by *intrinsic* markers (a,b) and imbedded in Euclidean three space (x^1,x^2,x^3). Any point on the surface may also be labeled with x^i, and we can form three independent Jacobians, $\partial(x^i,x^j)/\partial(a,b)$. Show that these form the components of an antisymmetric second-rank tensor. We can then define

$$dx^i \wedge dx^j = \frac{\partial(x^i,x^j)}{\partial(a,b)}\,da\,db$$

which is clearly a second-rank antisymmetric tensor. Illustrate that $ds^k = \frac{1}{2}\epsilon^{kij}\,dx^i \wedge dx^j$ represents an element of surface by considering some special cases.

3.2 Discuss how the above concepts can be generalized to surfaces of any number of dimensions in any Riemann space.

Bibliography

Compare Bibliography for Chap. 2. See also:

Fermi, E. (1922): Sopra i fenomeni che avvengono in vicinanza di una linea oraria, *Atti R. Accad. Lincei Rend. Cl. Sci. Fis. Mat. Nat.*, vol. 31_1, pp. 21–51.
Flanders, H. (1963): "Differential Forms," New York.
Grassmann, H. (1878): "Die Ausdehnungslehre von 1844," Leipzig.
Hammermesh, M. (1962): "Group Theory and Its Application to Physical Problems," New York.
Levi-Civita, T. (1927): "The Absolute Differential Calculus," London.
McConnell, A. J. (1936): "Applications of the Absolute Differential Calculus," London.
Murnaghan, F. D. (1922): "Vector Analysis and the Theory of Relativity," Baltimore.
Pirani, F. A. E. (1957): Tetrad formulation of general relativity theory, *Bull. Acad. Polon. Sci.*, vol. 5, pp. 143–147.
Schouten, J. A. (1951): "Tensor Analysis for Physicists," Oxford.
Schouten, J. A. (1954): "Ricci Calculus," 2d ed., Berlin-Göttingen-Heidelberg.
Synge, J. L. (1960): "Relativity: The General Theory," Amsterdam.
Walker, A. G. (1932): Relative coordinates, *Proc. Roy. Soc. Edinburgh*, vol. 52, p. 345.
Weinberg, S. (1972): "Gravitation and Cosmology," New York (sec. 13.1 on Killing vectors).

CHAPTER

4

Tensors in Physics

We have now reached the point where our mathematical machinery is sufficiently well developed to allow us to begin applying it to physics. Our purpose in this chapter is twofold: First, by applying the techniques of tensor analysis to familiar areas of physics, we shall gain facility in the use of our mathematical tools and gain insight as to how the relativistic gravitational theory might be formulated. Second, and much more important, we shall be led to answer several fundamental questions concerning the background concepts of the gravitational theory. We shall, for instance, demonstrate the suitability of the Lorentz metric for describing electromagnetic phenomena in mass-free space; we shall discuss the nature of the relation between the line element ds and the proper-time interval $d\tau$, and we shall choose and justify a tensor equation of motion. In the process of the investigation, it is also hoped that some equations of classical physics will be considerably simplified.

4.1 Maxwell's Equations in Tensor Form

We assume the reader is familiar with Maxwell's equations and classical electrodynamics as well as a reasonable amount of special relativity. The task of this section will be to formulate Maxwell's equations in covariant tensor form—one which is valid in all Riemannian coordinate systems. To do this we begin by considering the equations in just *one coordinate system at rest* and make a purely *formal* change to *tensor notation*. Afterwards, we can consider the transformation to other systems, such as those in motion relative to the original one; but in what follows it should always be kept in mind that we deal only with a single rest system until we explicitly state otherwise.

In résumé of classical electrodynamics we list the Maxwell equations in two conveniently separated pairs. The first pair,

$$\nabla \times \mathbf{H} - \frac{1}{c}\dot{\mathbf{E}} = \frac{1}{c}\mathbf{j} \qquad \dot{\mathbf{E}} \equiv \frac{\partial \mathbf{E}}{\partial t} \tag{4.1}$$

$$\nabla \cdot \mathbf{E} = \rho \qquad \nabla \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \tag{4.2}$$

(in Heaviside-Lorentz units) involves the current density $\mathbf{j}$ and the charge density ρ, which are the sources of the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$. We shall accordingly refer to this pair as the *source equations*. The second pair,

$$\nabla \times \mathbf{E} + \frac{1}{c}\dot{\mathbf{H}} = 0 \tag{4.3}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{4.4}$$

is homogeneous and deals only with the relation between the electric and the magnetic fields. Since the components of the electric and magnetic vector fields will be seen to be also the components of an electromagnetic field tensor in what follows, we can say that (4.3) and (4.4) are internal relations between the components of a single tensor. Accordingly, these will be referred to as the *internal equations*.

The sources of the field, the ρ and $\mathbf{j}$ of the source equations, cannot be independently specified. In general, (4.1) and (4.2) are not compatible unless a consistency relation between the sources is satisfied. We can obtain this consistency relation by taking the divergence of both sides of (4.1):

$$\nabla \cdot \nabla \times \mathbf{H} - \frac{1}{c}\nabla \cdot \dot{\mathbf{E}} = \frac{1}{c}\nabla \cdot \mathbf{j} \tag{4.5}$$

and inserting

$$\dot{\rho} = \nabla \cdot \dot{\mathbf{E}} \tag{4.6}$$

which is obtained by differentiating (4.2). Noting that the divergence of a curl is identically zero for any vector, we then find that

$$\dot{\rho} + \nabla \cdot \mathbf{j} = 0 \tag{4.7}$$

We have here a differential relation between charge density and current density which follows directly from the Maxwell equations. Unless this relation between the sources is satisfied, the Maxwell equations can have no solution, so we shall always assume that, for all physical sources, (4.7) is satisfied.

If we interpret $\mathbf{j}$ as the convection current

$$\mathbf{j} = \rho \mathbf{v} \tag{4.8}$$

where $\mathbf{v}$ is the velocity field of the material with charge density ρ, then (4.7) becomes identical with the continuity equation of fluid mechanics.

$$\dot{\rho} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{4.9}$$

This equation states that the quantity having density ρ is neither created nor destroyed; i.e., it is conserved. Thus we can interpret (4.7) as the *physical law of conservation of charge* as well as the *mathematical requirement for consistency of the source equations.*

Our first step in writing the Maxwell equations (4.1) to (4.4) and the conservation equation (4.7) with tensor notation is to choose a convenient set of four-dimensional coordinates and an appropriate metric tensor in this coordinate system. Following the lead of the special theory of relativity we shall utilize the space and time coordinates

$$x^0 = ct \qquad x^1 = x \qquad x^2 = y \qquad x^3 = z \tag{4.10}$$

where x, y, z are Cartesian space coordinates, and t is a time coordinate.

It will prove most convenient to adopt the convention, in the following chapters, that Greek indices run from 0 to 3 and Latin indices run from 1 to 3:

$$\begin{pmatrix} \nu = 0, 1, 2, 3 & \text{(Greek)} \\ j = 1, 2, 3 & \text{(Latin)} \end{pmatrix}$$

The question of what is an "appropriate" metric tensor to use with these coordinates can be approached in several different ways. By an appropriate metric tensor we here mean one which allows us—using the coordinates of special relativity—to write Maxwell's equations in a simple and elegant form which is easily generalized to a truly covariant form. The authors take the viewpoint that the motivation for the choice of an appropriate metric tensor should come from the elegant theory of classical electrodynamics and the structure of the Maxwell equations alone; if at all possible, it should not be necessary to depend on other theories such as special relativistic mechanics. Thus, in the following paragraphs, we shall attempt to obtain a quadratic form which plays a distinguished role in the Maxwell equations and which may therefore be adopted as the quadratic form of the metric tensor. [If the reader is not interested in this particular motivational approach to the choice of a metric tensor, he may skip directly to Eq. (4.43).]

To begin our search for a distinguished quadratic form we shall show that, associated with the Maxwell equations in vacuum, there is a class of particularly interesting surfaces. To be precise, let us ask the following question: Do there exist unique three-dimensional hypersurfaces, imbedded in the four-dimensional space we have chosen, on which the first derivatives of the **E** and **H** fields can be *discontinuous?* (In the region of the hypersurface we still demand that Maxwell's equations be satisfied.) If so, what is the nature of these surfaces? Such surfaces are quite important in the study of partial differential equations; they are generally referred to as characteristic hypersurfaces, or simply as characteristics. In the case of the first-order Maxwell equations, the characteristic has been defined as a hypersurface over which a discontinuity in the first derivatives can occur, but in the general case of an nth-order equation, the characteristic is defined as a hypersurface over which a discontinuity in the nth derivatives can occur.

In addition to being mathematically interesting, the notion of a characteristic can be seen to be physically meaningful. Suppose one passes a sharp pulselike electromagnetic disturbance through empty space. The front of the pulse can be very sharp if discontinuous first derivatives in the **E** and **H** fields are allowable on the hypersurface corresponding to the wave front. Thus we expect the characteristic hypersurfaces to correspond to allowed physical wave fronts. We know that the ideas of special relativity rest heavily on the assumption that such wave fronts propagate with the constant velocity c, so the motivation for investigating the characteristic hypersurfaces is apparent.

In vacuum the densities ρ and $\mathbf{j}$ are zero and Maxwell's equations accordingly become

$$\nabla \times \mathbf{H} - \frac{1}{c}\dot{\mathbf{E}} = 0 \tag{4.11}$$

$$\nabla \cdot \mathbf{E} = 0 \tag{4.12}$$

$$\nabla \times \mathbf{E} + \frac{1}{c}\dot{\mathbf{H}} = 0 \tag{4.13}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{4.14}$$

The equation of any three-dimensional hypersurface imbedded in a four-dimensional space can be written in the form

$$w(x^0,x^1,x^2,x^3) = 0 \tag{4.15}$$

We shall suppose that w has continuous first derivatives in all the variables, which means that it is a *smooth hypersurface.* If we further assume that $\partial w/\partial x^0$ is nonzero, we can solve (4.15) for x^0 as a function of x^1, x^2, and x^3 and write the equation of the three-dimensional hypersurface as

$$x^0 = h(x^1,x^2,x^3) \tag{4.16}$$

Then, on the hypersurface defined by (4.15), which we shall refer to as S, we have

$$w(x^0,x^1,x^2,x^3) = h(x^1,x^2,x^3) - x^0 = 0 \tag{4.17}$$

The electric and magnetic fields on S are functions of only x^1, x^2, x^3, since x^0 is specified by (4.17) when these coordinates are given. We shall therefore denote the electric and magnetic fields on S as the following vector functions of these three variables:

$$\hat{\mathbf{E}}(x^1,x^2,x^3) = \mathbf{E}(h,x^1,x^2,x^3) \tag{4.18}$$

$$\hat{\mathbf{H}}(x^1,x^2,x^3) = \mathbf{H}(h,x^1,x^2,x^3) \tag{4.19}$$

where h is the value of x^0 given by (4.17). The vector functions $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$ are assumed to have continuous first derivatives. Using the above definitions and a bit of vector algebra, we shall be able to obtain a pair of very useful relations (V. Fock, 1959) which the functions $\dot{\mathbf{E}}$, $\dot{\mathbf{H}}$, $\hat{\mathbf{E}}$, $\hat{\mathbf{H}}$, and h must obey on S. These relations will be the key to obtaining the characteristic surfaces of Maxwell's equations.

From (4.18), (4.19), and (4.17) we have

$$\frac{\partial \hat{E}_k}{\partial x^i} = \frac{\partial E_k}{\partial x^i} + \frac{\partial E_k}{\partial x^0}\frac{\partial h}{\partial x^i} \tag{4.20}$$

Setting $k = i$ and summing from 1 to 3, we obtain

$$\nabla \cdot \hat{\mathbf{E}} = \nabla \cdot \mathbf{E} + \frac{1}{c}\dot{\mathbf{E}} \cdot \nabla h \tag{4.21}$$

From (4.12) this gives

(4.22) $$\nabla \cdot \hat{\mathbf{E}} = \frac{1}{c}\dot{\mathbf{E}} \cdot \nabla h$$

For the magnetic field we may obtain the analogous result

(4.23) $$\nabla \cdot \hat{\mathbf{H}} = \frac{1}{c}\dot{\mathbf{H}} \cdot \nabla h$$

From (4.20) we also obtain

$$\frac{\partial E_k}{\partial x^i} - \frac{\partial E_i}{\partial x^k} + \frac{\partial E_k}{\partial x^0}\frac{\partial h}{\partial x^i} - \frac{\partial E_i}{\partial x^0}\frac{\partial h}{\partial x^k} = \frac{\partial \hat{E}_k}{\partial x^i} - \frac{\partial \hat{E}_i}{\partial x^k}$$

that is, in vector notation,

(4.24) $$\nabla \times \mathbf{E} + \frac{1}{c}\nabla h \times \dot{\mathbf{E}} = \nabla \times \hat{\mathbf{E}}$$

and as the analogous result for the magnetic field, we have

(4.25) $$\nabla \times \mathbf{H} + \frac{1}{c}\nabla h \times \dot{\mathbf{H}} = \nabla \times \hat{\mathbf{H}}$$

Substituting into these last two equations from Maxwell's equations in vacuum, (4.11) and (4.13), we obtain

(4.26) $$-\frac{1}{c}\dot{\mathbf{H}} + \frac{1}{c}\nabla h \times \dot{\mathbf{E}} = \nabla \times \hat{\mathbf{E}}$$

(4.27) $$\frac{1}{c}\dot{\mathbf{E}} + \frac{1}{c}\nabla h \times \dot{\mathbf{H}} = \nabla \times \hat{\mathbf{H}}$$

The scalar product of (4.26) and (4.27) with ∇h gives

(4.28) $$-\frac{1}{c}\nabla h \cdot \dot{\mathbf{H}} + \frac{1}{c}\nabla h \cdot \nabla h \times \dot{\mathbf{E}} = -\frac{1}{c}\nabla h \cdot \dot{\mathbf{H}} = \nabla h \cdot \nabla \times \hat{\mathbf{E}}$$

(4.29) $$\frac{1}{c}\nabla h \cdot \dot{\mathbf{E}} + \frac{1}{c}\nabla h \cdot \nabla h \times \dot{\mathbf{H}} = \frac{1}{c}\nabla h \cdot \dot{\mathbf{E}} = \nabla h \cdot \nabla \times \hat{\mathbf{H}}$$

Also the vector product of (4.26) and (4.27) with ∇h gives

(4.30) $$-\frac{1}{c}(\nabla h \times \dot{\mathbf{H}}) + \frac{1}{c}\nabla h \times (\nabla h \times \dot{\mathbf{E}}) = \nabla h \times (\nabla \times \hat{\mathbf{E}})$$

(4.31) $$\frac{1}{c}(\nabla h \times \dot{\mathbf{E}}) + \frac{1}{c}\nabla h \times (\nabla h \times \dot{\mathbf{H}}) = \nabla h \times (\nabla \times \hat{\mathbf{H}})$$

Expanding the double cross product and substituting from (4.26) and (4.27), we have

(4.32) $$\frac{1}{c}\dot{\mathbf{E}} - \nabla \times \hat{\mathbf{H}} + \frac{1}{c}\nabla h(\nabla h \cdot \dot{\mathbf{E}}) - \frac{1}{c}\dot{\mathbf{E}}(\nabla h)^2 = \nabla h \times (\nabla \times \hat{\mathbf{E}})$$

(4.33) $$\frac{1}{c}\dot{\mathbf{H}} + \nabla \times \hat{\mathbf{E}} + \frac{1}{c}\nabla h(\nabla h \cdot \dot{\mathbf{H}}) - \frac{1}{c}\dot{\mathbf{H}}(\nabla h)^2 = \nabla h \times (\nabla \times \hat{\mathbf{H}})$$

Finally, substituting from (4.28) and (4.29), we get

(4.34) $$\frac{1}{c}\dot{\mathbf{E}} - \nabla \times \hat{\mathbf{H}} + \nabla h(\nabla h \cdot \nabla \times \hat{\mathbf{H}}) - \frac{1}{c}\dot{\mathbf{E}}(\nabla h)^2 = \nabla h \times (\nabla \times \hat{\mathbf{E}})$$

(4.35) $$\frac{1}{c}\dot{\mathbf{H}} + \nabla \times \hat{\mathbf{E}} - \nabla h(\nabla h \cdot \nabla \times \hat{\mathbf{E}}) - \frac{1}{c}\dot{\mathbf{H}}(\nabla h)^2 = \nabla h \times (\nabla \times \hat{\mathbf{H}})$$

Rearrangement now gives the two key relations that we have been working toward:

(4.36) $$\frac{1}{c}(1 - [\nabla h]^2)\dot{\mathbf{E}} = \nabla \times \hat{\mathbf{H}} - \nabla h(\nabla h \cdot \nabla \times \hat{\mathbf{H}}) + \nabla h \times (\nabla \times \hat{\mathbf{E}})$$

(4.37) $$\frac{1}{c}(1 - [\nabla h]^2)\dot{\mathbf{H}} = -\nabla \times \hat{\mathbf{E}} + \nabla h(\nabla h \cdot \nabla \times \hat{\mathbf{E}}) + \nabla h \times (\nabla \times \hat{\mathbf{H}})$$

Two situations are now possible: Either the factor $1 - (\nabla h)^2$ which appears on the left side of (4.36) and (4.37) is zero or it is nonzero. Consider first the case where it is nonzero. We can then solve for $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ in terms of the continuous first derivatives of h, $\hat{\mathbf{E}}$, and $\hat{\mathbf{H}}$, so $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ are themselves continuous on S. We have thus determined the values $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ on S which represent, in view of (4.17), the normal derivatives of the vector fields $\mathbf{E}$ and $\mathbf{H}$ on the hypersurface S. By the well-known theory of the initial-value problem for first-order differential systems, the solution fields $\mathbf{E}$ and $\mathbf{H}$ are therefore uniquely determined as continuously differentiable functions of all four variables in a neighborhood of the hypersurface S. Thus, if $1 - (\nabla h)^2$ is not zero, the first derivatives of

$\mathbf{E}$ and $\mathbf{H}$ must be continuous across S, and S cannot be a characteristic hypersurface.

In the second case, where $1 - (\nabla h)^2$ is zero, we cannot solve for $\dot{\mathbf{H}}$ and $\dot{\mathbf{E}}$, which therefore remain undefined on S. This is the only situation for which the first derivatives of $\mathbf{E}$ and $\mathbf{H}$ can be discontinuous across S, so we obtain the condition that, for S to be a characteristic, $1 - (\nabla h)^2$ must vanish. Note, furthermore, that if $1 - (\nabla h)^2$ vanishes, the equations (4.36) and (4.37) then provide a relation between $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$; these functions cannot be arbitrarily prescribed on S. This is indeed a familiar property of characteristic surfaces: On a characteristic the fields must obey restrictive relations. (We shall see a further example of this property in Chap. 7, when we investigate the characteristic surfaces of Einstein's gravitational field equations.)

We now have the equation of the characteristics in the form

$$(\nabla h)^2 = 1 \tag{4.38}$$

which, by virtue of (4.17), is equivalent to

$$\left(\frac{\partial w}{\partial x^0}\right)^2 - (\nabla w)^2 = 0 \tag{4.39}$$

The reader may check that this is satisfied by the following class of hypersurfaces:

$$w = (x^0 - a^0)^2 - (x^1 - a^1)^2 - (x^2 - a^2)^2 - (x^3 - a^3)^2 = 0 \tag{4.40}$$

where the a's are arbitrary parameters. Indeed, one can also consider the a's to be functions of a single parameter λ. A continuous or discrete superposition of solutions of the form (4.40) is then easily made, and one can show that the envelopes of such superpositions constitute all the continuous and differentiable solutions of (4.39). In the theory of partial differential equations, a particular solution with the above outlined property is termed a *complete integral*. In the case of Maxwell's equations the complete integral (4.40) is the most important particular solution to (4.39).

By translating the origin of the coordinate system we can write the complete integral (4.40) as

$$(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = c^2t^2 - (x^2 + y^2 + z^2) = 0 \tag{4.41}$$

which clearly represents a three-dimensional sphere expanding at velocity c, or the four-dimensional *light cone*, which is familiar from special relativity. (In Chap. 8 characteristic surfaces will be discussed further.)

In (4.41) we have the equation of the most important characteristic written as a null quadratic form. In matrix notation it takes the form

$$(x^0x^1x^2x^3)\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}\begin{pmatrix}x^0\\ x^1\\ x^2\\ x^3\end{pmatrix} = 0 \tag{4.42}$$

We have here a quadratic form and its associated matrix which plays a very distinguished role in the propagation of electromagnetic wave fronts. On the other hand, the metric tensor is the associated matrix of another quadratic form, the line element, which also plays a fundamental role in the structure of a Riemann space. Thus, unless the matrix appearing in (4.42) and the metric tensor are the same, we are faced with the rather unsatisfactory situation of possessing two unrelated fundamental matrices. To avoid this situation and maintain the maximum amount of elegance and economy, we tentatively adopt the matrix appearing in (4.42) as the metric tensor in the ct, x, y, z coordinate system. The line element and the metric tensor are thus taken to be

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \qquad g_{\mu\nu} = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{pmatrix} \tag{4.43}$$

These are the familiar line element and metric tensor of special relativity which one usually obtains by quite different arguments. But let us note once again that the choice of this metric was here motivated directly by arguments based on the mathematical structure of the Maxwell equations, which is completely independent of all mechanical concepts.

It is now our task to show that the above metric is indeed a wise choice. To do this we shall rephrase Maxwell's equations in formal tensor notation. The elegance of the resultant equations will serve as justification for the choice. Further justification will also appear in Secs. 4.2 and 4.3, when we consider the consequences of also using this metric for the description of mechanical systems.

Let us begin the formal rephrasing of Maxwell's equations by defining a source four-vector.

$$s^\alpha = \left(\rho, \frac{1}{c}j_x, \frac{1}{c}j_y, \frac{1}{c}j_z\right) = \left(\rho, \frac{1}{c}\mathbf{j}\right) \tag{4.44}$$

The conservation of charge, Eq. (4.7), can be written using tensor notation as

$$c\rho_{|0} + j^i{}_{|i} = 0 \tag{4.45}$$

and in terms of the source vector s^α as

$$s^\alpha{}_{|\alpha} = 0 \tag{4.46}$$

Since the Christoffel symbols are zero in the coordinate system we are now using, we can just as well use the notation for *covariant* differentiation as ordinary differentiation. This switch to covariant differentiation is also advantageous if we wish to work in polar coordinates or in various other spatial curvilinear coordinate systems instead of rectangular coordinates, for then the equation needs no modification in the curvilinear system. Thus we write

$$s^\alpha{}_{\|\alpha} = 0 \tag{4.47}$$

When we later consider the transformation of the equations to a completely general coordinate system (in arbitrary motion), the use of covariant instead of ordinary differentiation will be of the utmost utility, as the reader has no doubt anticipated.

To write the source equations (4.1) and (4.2) in tensor notation, we define a 4×4 antisymmetric matrix composed of the components of the electric and magnetic fields in the following manner:

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & H_z & -H_y \\ -E_y & -H_z & 0 & H_x \\ -E_z & H_y & -H_x & 0 \end{pmatrix} \tag{4.48}$$

This is the Minkowski *electromagnetic field tensor*. (At this point, of course, we are not yet justified in calling it a tensor, but only a matrix.) Using this matrix we can write the source equations as

$$F^{\mu\nu}{}_{|\nu} = s^\mu \qquad F^{\mu\nu}{}_{\|\nu} = s^\mu \tag{4.49}$$

where we have again used covariant differentiation instead of ordinary differentiation, for the same reasons as before. The easiest method of verifying that (4.49) is equivalent to the two Maxwell equations (4.1) and (4.2) is simply to substitute appropriate values of μ and ν and check that (4.1) and (4.2) result. For $\mu = 0$, we have

$$F^{00}{}_{|0} + F^{01}{}_{|1} + F^{02}{}_{|2} + F^{03}{}_{|3} = s^0 \tag{4.50}$$

or equivalently,

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \rho \tag{4.51}$$

which is precisely the Maxwell equation (4.2). For $\mu = 1$, we have

$$F^{10}{}_{|0} + F^{11}{}_{|1} + F^{12}{}_{|2} + F^{13}{}_{|3} = s^1 \tag{4.52}$$

or equivalently,

$$-\frac{1}{c}\dot{E}_x + \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \frac{1}{c}j_x \tag{4.53}$$

which is the x component of the Maxwell equation (4.1). The other components of (4.1) may be similarly checked.

In like manner the reader may satisfy himself that in tensor notation the internal Maxwell equations (4.3) and (4.4) are equivalent to

$$F_{\mu\nu|\lambda} + F_{\lambda\mu|\nu} + F_{\nu\lambda|\mu} = 0 \tag{4.54}$$

or, using the antisymmetrization notation introduced in Chap. 3,

$$\{F_{\mu\nu|\lambda}\} = 0 \tag{4.55a}$$

Here we need *not* replace ordinary by covariant differentiation since (4.55*a*) is already in covariant form, as we showed in Sec. 3.4. As we noted in Sec. 3.5, this is also equivalent to the dual tensor equation

$$*F^{\mu\nu}{}_{\|\nu} = 0 \tag{4.55b}$$

These formulas are best understood if we write out explicitly the dual matrix

$$*F_{\mu\nu} = \begin{pmatrix} 0 & H_x & H_y & H_z \\ -H_x & 0 & E_z & -E_y \\ -H_y & -E_z & 0 & E_x \\ -H_z & E_y & -E_x & 0 \end{pmatrix} \tag{4.56a}$$

which arises from $F_{\mu\nu}$ by interchange of **E** and **H**. If we construct the contravariant dual tensor, using $g^{\mu\nu} = g_{\mu\nu}$ in (4.42), we obtain

$$*F^{\mu\nu} = \begin{pmatrix} 0 & -H_x & -H_y & -H_z \\ H_x & 0 & E_z & -E_y \\ H_y & -E_z & 0 & E_x \\ H_z & E_y & -E_x & 0 \end{pmatrix} \tag{4.56b}$$

That is, $*F^{\mu\nu}$ differs from $F^{\mu\nu}$ by the change of **E** into $-$**H** and **H** into **E**. The differential expressions $(F^{i\nu}{}_{|\nu}) = \nabla \times \mathbf{H} - (1/c)\dot{\mathbf{E}}$ therefore go over to $(*F^{i\nu}{}_{|\nu}) = \nabla \times \mathbf{E} + (1/c)\dot{\mathbf{H}}$.

Equation (4.55*a*) is, as we found in Sec. 3.5, simply the necessary and sufficient condition that $F_{\mu\nu}$ is closed and has a tensor potential. That is, there exists a four-vector ϕ_μ such that

$$F_{\mu\nu} = \phi_{\mu|\nu} - \phi_{\nu|\mu} \tag{4.57a}$$

Now, in classical electrodynamics, the only physically significant quantities are the electric and magnetic fields **E** and **H** which appear in the field tensor $F_{\mu\nu}$. Therefore the four-vector function ϕ_μ has no *direct* physical meaning; only its four-dimensional curl has physical meaning. It is thus clear that we may make a so-called *gauge transformation* on ϕ_μ; that is, we can add an arbitrary four-dimensional gradient $\Psi_{|\mu}$ to ϕ_μ, without altering $F_{\mu\nu}$ and therefore without altering the physical situation. That is,

$$(\phi_\mu + \Psi_{|\mu})_{|\nu} - (\phi_\nu + \Psi_{|\nu})_{|\mu} = \phi_{\mu|\nu} - \phi_{\nu|\mu} = F_{\mu\nu} \tag{4.57b}$$

We say, then, that the physically meaningful $F_{\mu\nu}$ tensor is *gauge-invariant;* i.e., it is not altered by a gauge transformation.

We have now obtained the Maxwell equations and their associated conservation law, all written formally in tensor notation but taken to be valid only in some specified rest system.

$$\begin{aligned} F^{\mu\nu}{}_{\|\nu} &= s^\mu \\ *F^{\mu\nu}{}_{\|\nu} &= 0 \qquad \text{or} \qquad \{F_{\mu\nu|\lambda}\} = 0 \\ s^\mu{}_{\|\mu} &= 0 \end{aligned} \tag{4.58}$$

In classical electrodynamics there is *no a priori rule for transforming the fields and the field equations* to a system in motion relative to the original rest frame. The rules proposed for the transformation previous to the application of relativity to electrodynamics were rather unconvincing. However, with the equations written in *tensor notation*, we are presented

with a more convincing solution to the problem. The temptation to postulate that the Maxwell equations in tensor notation are indeed *tensor equations* is very great, and we accordingly make this assumption. Then, under a general coordinate transformation, the *electric and magnetic fields transform as the indicated components of* $F_{\mu\nu}$, which is now a tensor, and the equations for the fields remain the same in all systems. The transformation properties of the fields which we postulate by this assumption can be tested in experiments involving moving material media and, indeed, agree quite well with experiment.

When Minkowski first introduced the $F_{\mu\nu}$ tensor into electrodynamics, he had in mind that it should transform as a tensor only under Lorentz transformations. However, as we see here, no such restriction appears necessary, for the Maxwell equations go over very easily indeed into a completely covariant form.

The reader should not be blinded by our mathematical transformations into assuming that the statement "$F^{\mu\nu}$ is a tensor" is a purely mathematical one. It is a very important and far-reaching physical principle which can be *motivated* by mathematical elegance, but must also be *tested* by physical experiment. The fact that $F^{\mu\nu}$ is a tensor under Lorentz transformation embodies a large part of special relativity theory. Our assumption that $F^{\mu\nu}$ is a tensor in a general Riemannian space-time leads to important consequences for the electrodynamics in accelerated systems of reference. The methodological principle that laws of nature which appear in tensor form in a particular coordinate system should be interpreted as valid in every system is called the *principle of covariance.* Its philosophical motivation is the postulate that no coordinate system should be distinguished in the formation of physical laws. It is, however, mathematically somewhat ambiguous and comes in practice to the old principle that we should try to explain facts with the simplest and most aesthetically satisfactory theory.

Now that we assume the Maxwell equations are tensor equations, we can investigate their form further. We note that the internal equations (4.55*a*) contain no terms involving the metric tensor. If the role of the metric tensor in the source equations (4.49) could be reduced to a minimum, these equations might be somewhat simplified. To see this let us write them out explicitly, working now in an arbitrary coordinate system:

$$F^{\mu\nu}{}_{\|\nu} = F^{\mu\nu}{}_{|\nu} + \begin{Bmatrix} \nu \\ \nu\ \alpha \end{Bmatrix} F^{\mu\alpha} + \begin{Bmatrix} \mu \\ \alpha\ \nu \end{Bmatrix} F^{\alpha\nu} = s^\mu \tag{4.59}$$

We found in (3.11) that the "contracted" Christoffel symbol can be written as

$$\begin{Bmatrix} \nu \\ \nu\ \alpha \end{Bmatrix} = \frac{1}{2g}\, g_{|\alpha} \tag{4.60}$$

where g is the determinant of the metric tensor. Next note that $F^{\alpha\nu}$ is antisymmetric while $\left\{ {\mu \atop \alpha\ \nu} \right\}$ is symmetric in α and ν. Thus the sum over α and ν of the product will cancel in pairs to give a zero result:

$$\left\{ {\mu \atop \alpha\ \nu} \right\} F^{\alpha\nu} = 0 \tag{4.61}$$

Therefore, (4.59) can be written

$$F^{\mu\nu}{}_{\|\nu} = F^{\mu\nu}{}_{|\nu} + \frac{1}{2g}(g_{|\alpha})F^{\mu\alpha} = s^\mu \tag{4.62}$$

or equivalently,

$$F^{\mu\nu}{}_{\|\nu} = F^{\mu\nu}{}_{|\nu} + \frac{1}{\sqrt{-g}}(\sqrt{-g})_{|\alpha}F^{\mu\alpha} = \frac{1}{\sqrt{-g}}(\sqrt{-g}\,F^{\mu\nu})_{|\nu} = s^\mu \tag{4.63}$$

In this form only the single quantity g enters the equation instead of the 16 components of $g_{\mu\nu}$ as in (4.49).

If we denote the antisymmetric tensor density (Sec. 3.5) associated with the $F^{\mu\nu}$ field tensor as

$$\mathfrak{F}^{\mu\nu} = \sqrt{-g}\,F^{\mu\nu} \tag{4.64}$$

and the density associated with s^μ as

$$\mathfrak{S}^\mu = \sqrt{-g}\,s^\mu \tag{4.65}$$

we can write (4.63) in another convenient form:

$$\mathfrak{F}^{\mu\nu}{}_{|\nu} = \mathfrak{S}^\mu \tag{4.66}$$

As an exercise to illustrate the great advantage of working with the Maxwell equations in the above tensor form, let us derive the charge conservation equation directly from (4.66). Differentiation with respect to x^μ gives

$$\mathfrak{F}^{\mu\nu}{}_{|\nu|\mu} = \mathfrak{S}^\mu{}_{|\mu} \tag{4.67}$$

Since $\mathfrak{F}^{\mu\nu}$ is antisymmetric in μ and ν while the differential operator $\partial^2/(\partial x^\nu \partial x^\mu)$ is symmetric, the sum on the left vanishes, and we obtain

$$\mathfrak{S}^\mu{}_{|\mu} = (\sqrt{-g}\,s^\mu)_{|\mu} = 0 \tag{4.68}$$

From (3.12) we have

$$s^\mu{}_{\|\mu} = \frac{1}{\sqrt{-g}}(\sqrt{-g}\,s^\mu)_{|\mu} = \frac{1}{\sqrt{-g}}\mathfrak{S}^\mu{}_{|\mu} \tag{4.69}$$

so since g is never zero, we obtain, from Eqs. (4.68) and (4.69), the conservation equation (4.47).

As a second illustration of the ease of working with the tensor form of Maxwell's equations, let us consider a "scale change" of the metric tensor: $g_{\mu\nu} \to A(x^\alpha)g_{\mu\nu}$, where $A(x^\alpha)$ is an arbitrary function of position. (It should be noted that such a scale change is completely unrelated to an ordinary transformation of coordinates.) We shall show that Maxwell's equations in free space are actually invariant under such a scale change. In order to show this, we write Maxwell's equations in the form displayed in (4.57*a*) and (4.66):

$$F_{\mu\nu} = \phi(x^\alpha)_{\mu|\nu} - \phi(x^\alpha)_{\nu|\mu} \qquad (\text{equivalently, } \{F_{\mu\nu|\lambda}\} = 0) \tag{4.70a}$$

$$\mathfrak{F}^{\mu\nu}{}_{|\nu} = (\sqrt{-g}\,F^{\mu\nu})_{|\nu} = 0 \tag{4.70b}$$

A scale change of the metric then replaces $g_{\mu\nu}$ by

$$\tilde{g}_{\mu\nu} = A g_{\mu\nu} \tag{4.71}$$

The tensor $\tilde{g}^{\mu\nu}$ is defined as the inverse of $\tilde{g}_{\mu\nu}$ and is therefore clearly given by

$$\tilde{g}^{\mu\nu} = \frac{1}{A}g^{\mu\nu} \tag{4.72}$$

Under the scale change defined by (4.71) and (4.72), the four-vector $\phi(x^\alpha)_\mu$ can be consistently considered (by definition) to remain unchanged since it is simply a function of position:

$$\tilde{\phi}(x^\alpha)_\mu = \phi(x^\alpha)_\mu \tag{4.73}$$

This implies that $F_{\mu\nu}$ is unchanged also:

$$\tilde{F}_{\mu\nu} = \tilde{\phi}_{\mu|\nu} - \tilde{\phi}_{\nu|\mu} = \phi_{\mu|\nu} - \phi_{\nu|\mu} = F_{\mu\nu} \tag{4.74}$$

and that equation (4.70*a*) is therefore invariant under a change of the metric scale.

To obtain the doubly contravariant tensor $\bar{F}^{\mu\nu}$ which appears in (4.70b), we must raise both indices of $\bar{F}_{\mu\nu}$ (4.74), using $\bar{g}^{\mu\nu}$ in (4.72). This gives

$$\bar{F}^{\mu\nu} = \bar{g}^{\mu\alpha}\bar{g}^{\nu\beta}\bar{F}_{\alpha\beta} = \frac{1}{A^2}g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta} = \frac{1}{A^2}F^{\mu\nu} \tag{4.75}$$

Combining this with the following expression for $\sqrt{-\bar{g}}$, which follows from the definition (4.71),

$$\sqrt{-\bar{g}} = \sqrt{-A^4 g} = A^2\sqrt{-g} \tag{4.76}$$

we see that

$$\sqrt{-\bar{g}}\,\bar{F}^{\mu\nu} = \sqrt{-g}\,F^{\mu\nu} \tag{4.77}$$

That is, the tensor density $\mathfrak{F}^{\mu\nu} = \sqrt{-g}\,F^{\mu\nu}$ is invariant; it then follows immediately that Eq. (4.70b) is also invariant under the scale change (4.71).

4.2 Proper-Time and the Equations of Motion via an Example in Relativistic Mechanics

In the previous section we dealt with the formulation of classical electrodynamics in a four-dimensional Riemann space; it did not prove necessary to consider questions of measurement concerning meter sticks and clocks and their relation to the four-dimensional Riemann space. Nor were we concerned with the paths of particles in the Riemann space, since we dealt only with tensor *fields*. The above are all specifically mechanical concepts, and, as we have noted, the results of the preceding section were purposely obtained in a way which was independent of mechanical notions. In this section we shall consider these mechanical questions, which we have so far carefully avoided.

In the hope of keeping the present development well grounded in familiar physical concepts, we shall consider the specific example of a coordinate system rotating at constant angular velocity and investigate the "fictitious" Coriolis and centrifugal forces associated with the rotation. Reasons for considering this example are twofold: First, we know that the equivalence principle asserts that the "fictitious" force due to acceleration of the coordinate system and the "real" force of gravity are in essence the same sort of phenomena. By considering these simple examples of fictitious forces we can hope to learn something of how forces in general occur in the context of relativity theory. Second, the investigation of this example will lead naturally to a reasonable solution of the general mechanical questions we cited above—the nature of space-time measurements and the equations of motion for particles in a Riemann space.

We begin as in the previous section with an inertial system and the coordinates of special relativity, ct, x, y, z. Because of the success of using the Lorentz metric to formulate the free-space Maxwell equations in a four-dimensional Riemann space, we shall tentatively carry it over (in the spirit of elegance and economy) to use also in describing mechanical phenomena in a mass-free Riemann space. The line element and metric are then assumed to be those in (4.43):

$$\begin{aligned} ds^2 &= (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \\ &= c^2\,dt^2 - (dx^2 + dy^2 + dz^2) \end{aligned} \tag{4.78}$$

$$g_{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}$$

This particular metric tensor will from now on be called the Lorentz metric tensor, and the corresponding metric, the Lorentz metric.

It will be convenient in this section to work with cylindrical coordinates instead of the Cartesian coordinates used in (4.78). In terms of cylindrical coordinates r, $\bar{\varphi}$, and z, we have

$$x = r\cos\bar{\varphi} \qquad y = r\sin\bar{\varphi} \qquad z = z \tag{4.79}$$

and

$$dx^2 + dy^2 + dz^2 = dr^2 + r^2\,d\bar{\varphi}^2 + dz^2 \tag{4.80}$$

Thus the Lorentz metric in cylindrical coordinates is

$$ds^2 = c^2\,dt^2 - (dr^2 + r^2\,d\bar{\varphi}^2 + dz^2) \tag{4.81}$$

We now define a transformation to a new t, r, φ, z system rotating about the z axis of the above inertial system with angular velocity w. To visualize the situation, one may think of the new system as being attached to a material disk which rotates with respect to the original coordinate system. This involves a change in $\bar{\varphi}$ only, which is clearly

$$\varphi = \bar{\varphi} - wt \tag{4.82}$$

The line element in this rotating system is easily obtained from (4.81) by the use of (4.82):

$$\begin{aligned} ds^2 &= c^2\,dt^2 - [dr^2 + r^2\,d\varphi^2 + 2wr^2\,d\varphi\,dt + w^2r^2\,dt^2 + dz^2] \\ &= (c^2 - w^2r^2)\,dt^2 - (dr^2 + r^2\,d\varphi^2 + 2wr^2\,d\varphi\,dt + dz^2) \end{aligned} \tag{4.83}$$

At this point we have obtained "complete" knowledge of the abstract geometry of space in the rotating coordinate system in the sense that the metric and the line element are known functions of the coordinates. We can at present go no further toward our goal of a description of physical measurements and mechanical processes in the rotating system, since we as yet have no way of relating the geometry of the space to clocks, measuring rods, and the motion of particles. To continue we must *interpret* the abstract geometry (embodied in the functional form of the line element) in mechanistic physical terms.

To interpret the abstract four-dimensional geometry and link it to reality by identifying a physical measurement with the evaluation of a geometrical object, we first need to define a few geometrical terms which one uses in four-space. An event is a point in four-space: a world-point. A single infinity of events forms a curve which we shall call a world-line or history if we are dealing with the representative point of a physical particle. The arc length between two events along a world-line is a geometrical invariant.

Let us now investigate what we mean by a measured proper-time interval in the special theory of relativity and obtain its relation to the line element ds^2. Suppose the line element between two events is $ds^2 = c^2\,dt^2 - (dx^2 + dy^2 + dz^2) \geq 0$, that is, a timelike interval. If we choose the Lorentz system in which the three-dimensional separation between the events is zero (the *proper* system), then

$$ds^2 = c^2\,dt^2 \tag{4.84}$$

Choosing the positive root, we then have

$$dt = \frac{ds}{c} \tag{4.85}$$

The interval ds/c is termed the proper-time interval between the events and corresponds, by the above comments, to the time interval that would be measured by a physicist to whom both events occurred at the same point in a three-dimensional frame to which he is attached. In practice this corresponds, for instance, to a physicist measuring the proper lifetime of a μ-meson by riding with it from the event corresponding to the creation of the meson to the event corresponding to its decay. Now, in general relativity, we wish to define the proper-time interval between two infinitesimally close events in some *invariant* manner which agrees with the above expression for the proper-time interval in the special case of the Lorentz metric. Thus it is very reasonable to make the following definition: In general relativity, an infinitesimal proper-time interval between two neighboring events is defined as the invariant generalization of (4.85) to an arbitrary Riemann space:

$$d\tau \equiv \frac{ds}{c} = \frac{1}{c}\sqrt{g_{00}}\,dx^0 \qquad \text{where } dx^\mu = 0 \text{ for } \mu \neq 0 \tag{4.86}$$

Notice that this definition of proper-time interval has been given only for infinitesimals and that it is physically meaningful only for $ds^2 \geq 0$, since otherwise $d\tau$ would be imaginary.

To illustrate our definition of proper-time intervals, let us investigate the relation of the proper-time interval $d\tau$ to the coordinate-time interval dt for two events which occur with the same spatial coordinates r, φ, z in the rotating coordinate system in our example above. We set the coordinate intervals dr, $d\varphi$, and dz equal to zero in (4.83) and obtain

$$ds^2 = (c^2 - w^2r^2)\,dt^2 \tag{4.87}$$

Thus, by the proper-time definition (4.86),

$$d\tau = \frac{ds}{c} = \left(1 - \frac{w^2r^2}{c^2}\right)^{1/2} dt \tag{4.88}$$

This can be interpreted as follows: $d\tau$ is the time interval between the events as measured by an observer attached to the rotating disk. On the other hand, dt is the time interval as measured by an observer attached to the nonrotating coordinate system and who uses the standard coordinate time t. How the physical measurements of time are carried out in both cases will be elucidated when we talk of finite time intervals a few paragraphs later. Thus (4.88) tells us that the time interval between the two events is different for these two observers, and indeed the ratio is $\sqrt{1 - w^2r^2/c^2}$. We can easily verify that this relation, which we obtained by an interpretation of the abstract geometry of the rotating system, is in agreement with the notions of special relativity. The linear velocity of a fixed point in the rotating coordinate system is given by

$v = rw$, so we can write the proper-time relation as

$$d\tau = \left(1 - \frac{v^2}{c^2}\right)^{1/2} dt \tag{4.89}$$

which we recognize as the familiar Lorentz time-dilation equation. It should be noted that this interpretation places an upper limit on the useful range of values of r; unless r is less than c/w, a fixed point in the rotating coordinate system will exceed the velocity of light, making that system inaccessible to a physical observer. [Of course as a formal mathematical operation the transformation (4.82) may still be made.]

Let us next apply (4.83) to measure three-dimensional space intervals. The cross term $-2wr^2\,d\varphi\,dt$ presents here a mathematical difficulty. We are used to line elements ds^2, which are quadratic in all differentials of our space geometry, and the cross term violates this rule. We must thus introduce a new time differential in the rotating system which removes this term. We are forced to a redefinition of simultaneity:

$$dt^* = dt - \frac{wr^2}{c^2 - w^2r^2}\,d\varphi \tag{4.90}$$

The new choice of the time differential brings (4.83) into the form

$$ds^2 = (c^2 - w^2r^2)\,dt^{*2} - \left(dr^2 + \frac{c^2}{c^2 - w^2r^2}\,r^2\,d\varphi^2 + dz^2\right) \tag{4.91}$$

The intervals dr, $d\varphi$, and dz corresponding to simultaneity according to $dt^* = 0$ are the intervals which a physicist would measure on the rotating frame. We read off from (4.91) that, with this understanding, a correct longitudinal Lorentz dilatation takes place, while scales transverse to the local velocity remain unchanged. In general, a moving observer will try to define his local space-time intervals in such a way that the line element is *locally* Lorentzian, as in (4.91). Observe, however, that the differential (4.90) is not exact; that is, we cannot introduce markers t^*, r, φ, z *in the large* which would lead to the line element (4.91).

Let us investigate further the general properties of proper-time intervals and coordinate-time intervals and the relative usefulness of the two concepts. We now extend the definition of proper-time to finite intervals; a proper-time interval is defined to be the invariant proportional (with coefficient $1/c$) to the arc length along a world-line and therefore has the very desirable property that it is independent of any reference frame. It will be useful in formulating basic physical laws and in describing fundamental processes if we identify it, considered as a geometrical object, with a physically measurable time interval. This is always implicitly done by relativists when considering that physical time is measured by the ticking of an atomic oscillator and that the period of such an oscillator is an invariant that remains the same during the history of the atom considered (is independent of the atom's age and position in three-space). This identification amounts to a basic postulate in the theory of relativity. It has been called by Synge (Synge, 1956) the *chronometric hypothesis* and is stated most clearly using a four-dimensional picture: On the world-line of a material particle there exists or can be thought to exist a discrete set of events separated by equal proper-time intervals. These events can be created by conceiving a standard atomic clock carried by the material particle. Furthermore, this postulate can have meaning only if it does not depend on the type of atomic clock used; therefore one needs to make the consistency hypothesis that, for a fixed arc length of an arbitrary common world-line, the ratio of the number of ticks of two atomic clocks is a natural constant.

From the above paragraph it is clear that an infinitesimal proper-time interval is a very useful concept. However, let us now show that the notion of proper-time *in the large* as opposed to an infinitesimal proper-time interval meets with difficulty and is not such a useful notion. The total elapsed proper-time between widely separated events which we define as the integral

$$\int_{\text{event 1}}^{\text{event 2}} \frac{ds}{c} = \frac{s}{c} = \tau \tag{4.92}$$

is clearly dependent on the *path* of integration, i.e., on the world-line followed by the standard clock between events 1 and 2. Therefore, if we specify an initial zero point of proper-time at some space-time point x_0^μ, we cannot extend a set of proper-time values throughout space in a unique way and we cannot label each event with one proper-time value. We say, then, that proper-time is not integrable; we cannot uniquely extend a proper-time over the space-time manifold. This fact gives rise to the well-known twin paradox in the special theory of relativity. The paradoxical nature of an age difference between the twins disappears once one stops thinking in terms of Newtonian pictures; following our definition of proper-time (age), one has to compare the arc lengths of the two different world-lines (histories) of the twins in a four-dimensional diagram. These two lengths have no reason to be equal.

The situation with the coordinate-time is in a sense opposite to that with proper-time. The coordinate-time interval between events is not an invariant, but by the definition of a Riemann space, each space-time point possesses a unique coordinate-time label. Thus the coordinate-time has unambiguous meaning at each point in space-time, and is hence an integrable quantity. It follows that, when we speak of *widely separated events*, the concept of coordinate-time is useful, for the coordinate-time separation between any two events—however widely separated—is unique and well defined in any given coordinate system. For example, if the period of an atomic oscillator is expressed in coordinate-time as dt by an observer at some point in space-time, it is the same coordinate-time interval dt to all observers throughout space-time. We may sum up: Proper-time intervals $d\tau$ are coordinate-invariant, but not integrable, whereas coordinate-time intervals dt are integrable, but not invariant.

The concepts of proper-time and coordinate-time which we have discussed in the foregoing paragraphs often prove to complement each other in practice in a very satisfying way. As an example, suppose we wish to compare "corresponding" proper-time intervals at different points in space. We can proceed as follows: At some initial point in space-time x_0^μ we specify the duration of a physical event in terms of an invariant, measurable proper-time interval $d\tau(x_0^\mu)$. The g_{00} element of the metric tensor then allows us to relate this to a *corresponding* coordinate-time interval dt by means of the defining relation

$$d\tau(x_0^\mu) = \sqrt{g_{00}(x_0^\mu)}\, dt \tag{4.93}$$

This dt, being an integrable coordinate-time interval, has unique meaning throughout space: to $d\tau(x_0^\mu)$ there corresponds only one value of dt which is the same throughout space. However, at a distant point x^μ this dt corresponds to quite another proper-time interval $d\tau(x^\mu)$, which is given by a relation analogous to (4.93):

$$d\tau(x^\mu) = \sqrt{g_{00}(x^\mu)}\, dt \tag{4.94}$$

Thus the two proper-time intervals $d\tau(x_0^\mu)$ and $d\tau(x^\mu)$, which have only local meaning but correspond to the same coordinate-time interval dt, are related to each other by

$$\frac{d\tau(x^\mu)}{d\tau(x_0^\mu)} = \left(\frac{g_{00}(x^\mu)}{g_{00}(x_0^\mu)}\right)^{1/2} \tag{4.95}$$

as follows directly from (4.93) and (4.94). This relation between corresponding proper-time intervals is the basis for calculations of the red shift, and we shall return to it in Sec. 4.4.

Let us now move on to the second basic problem of this section: the equations of motion for a particle in a four-dimensional Riemann space.

We showed in Chap. 2 that the force-free motion of a classical particle or a system of classical particles can be expressed in differential geometric form. Specifically, we showed that if x^i are the generalized coordinates of a system of n degrees of freedom, and if the kinetic energy of the system is of the form

$$T = \frac{1}{2}\sum_{\substack{i=1\\k=1}}^{n} g_{ik}\frac{dx^i}{dt}\frac{dx^k}{dt} \tag{4.96}$$

(where g_{ik} is independent of time), then the force-free motion of the system occurs along the extremal curves or configuration-space geodesics of the variational problem

$$\delta\int_{t_i}^{t_f} T\, dt = \delta\int_{t_i}^{t_f}\frac{1}{2}\sum_{\substack{i=1\\k=1}}^{n} g_{ik}\frac{dx^i}{dt}\frac{dx^k}{dt}\, dt \tag{4.97}$$

In terms of the line element,

$$ds^2 = g_{ik}\, dx^i\, dx^k \tag{4.98}$$

this is expressible in differential geometric form as

$$\delta\int_{t_i}^{t_f} T\, dt = \delta\int_{t_i}^{t_f}\frac{1}{2}\left(\frac{ds}{dt}\right)^2 dt = 0 \tag{4.99}$$

For the mechanical developments in Chap. 2, we treated time as an invariant scalar and not as a coordinate. In relativity theory, on the other hand, we wish to treat time as an additional coordinate and use a line element involving four coordinate intervals. To pursue this further, let us consider a line element very similar to the classical form (4.98) but involving also a time interval dt·

$$ds^2 = c^2\, dt^2 - g_{ik}\, dx^i\, dx^k \tag{4.100}$$

The line element corresponding to the Lorentz metric expressed in an inertial system is of precisely this form. We shall now show that the variational problem using this line element,

$$\delta\int_{s_i}^{s_f} ds = \delta\int_{s_i}^{s_f}\left[c^2\left(\frac{dt}{ds}\right)^2 - g_{ik}\frac{dx^i}{ds}\frac{dx^k}{ds}\right]^{1/2} ds = 0 \tag{4.101}$$

reduces to the nonrelativistic form (4.99) and therefore is an invariant representation of the force-free motion of a classical dynamical system. Since we are using s as parameter, we can just as well square the integrand (as we discussed in Sec. 2.3) and consider the equivalent variational problem

$$\delta \int \left\{c^2 \left(\frac{dt}{ds}\right)^2 - g_{ik} \frac{dx^i}{ds}\frac{dx^k}{ds}\right\} ds = 0 \tag{4.102}$$

The Euler-Lagrange equations for this variational problem are

$$\frac{\partial L}{\partial x^\mu} - \frac{d}{ds}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right) = 0 \tag{4.103}$$

where L is the integrand and $\dot{x}^\mu = dx^\mu/ds$. For $x^\mu = ct$, we have the equation

$$\frac{d}{ds}\left(2c \frac{dt}{ds}\right) = 0 \tag{4.104}$$

so that

$$\frac{dt}{ds} = \text{const} \tag{4.105}$$

By multiplying the coordinate time (which is just a marker) by an arbitrary constant we do not alter any physics, so we can choose the constant in Eq. (4.105) to be 1. This gives

$$dt = ds \tag{4.106}$$

allowing us to replace ds in the variational problem (4.101) by dt. The problem then becomes

$$\delta \int_{t_i}^{t_f} \left[c^2 - g_{ik}\frac{dx^i}{dt}\frac{dx^k}{dt}\right] dt = -\delta \int_{t_i}^{t_f} \left[g_{ik}\frac{dx^i}{dt}\frac{dx^k}{dt}\right] dt = 0 \tag{4.107}$$

which is, as we wished to show, the same as the nonrelativistic form (4.99).

Let us investigate the results of tentatively applying the geodesic equations as equations of motion in the example of the rotating coordinate system in which the line element is somewhat more complicated than (4.100). Using the form of the line element given in (4.83), the variational problem for the geodesics is

$$0 = \delta \int ds = \delta \int \left\{(c^2 - r^2w^2)\left(\frac{dt}{ds}\right)^2 - \left[\left(\frac{dr}{ds}\right)^2 + r^2\left(\frac{d\varphi}{ds}\right)^2 + 2wr^2 \frac{d\varphi}{ds}\frac{dt}{ds} + \left(\frac{dz}{ds}\right)^2\right]\right\}^{1/2} ds \tag{4.108}$$

If we again use s as a parameter, the variational problem can also be written in the equivalent form

$$0 = \delta \int \left\{(c^2 - r^2w^2)\left(\frac{dt}{ds}\right)^2 - \left[\left(\frac{dr}{ds}\right)^2 + r^2\left(\frac{d\varphi}{ds}\right)^2 + 2wr^2 \frac{d\varphi}{ds}\frac{dt}{ds} + \left(\frac{dz}{ds}\right)^2\right]\right\} ds \tag{4.109}$$

Accordingly, the Euler-Lagrange equation for $x^\mu = r$ is

$$-2rw^2\dot{t}^2 - 2r\dot{\varphi}^2 - 4wr\dot{\varphi}\dot{t} = -2\ddot{r} \tag{4.110}$$

Rearrangement gives

$$\ddot{r} = rw^2\dot{t}^2 + r\dot{\varphi}^2 + 2wr\dot{\varphi}\dot{t} \tag{4.111}$$

Similarly, the Euler-Lagrange equation for $x^\mu = z$ is

$$\ddot{z} = 0 \tag{4.112}$$

and that for $x^\mu = \varphi$ is

$$\frac{d}{ds}[r^2\dot{\varphi} + wr^2\dot{t}] = 0 \tag{4.113}$$

This last differential equation yields immediately

$$r^2\dot{\varphi} + wr^2\dot{t} = \text{const} \tag{4.114}$$

Finally, the Euler-Lagrange equation for $x^\mu = t$ is

$$\frac{d}{ds}[(c^2 - w^2r^2)\dot{t} - wr^2\dot{\varphi}] = 0 \tag{4.115}$$

which is also solved at once to give

$$(c^2 - r^2w^2)\dot{t} - wr^2\dot{\varphi} = \text{const} \tag{4.116}$$

These four Euler-Lagrange equations can now be shown to yield the familiar fictitious forces associated with a rotating system.

To show this, let us begin by multiplying (4.114) by the constant w:

$$wr^2\dot{\varphi} + w^2r^2\dot{t} = \text{const} \tag{4.117}$$

Adding this to (4.116), we have

$$c^2\dot{t} = \text{const} \tag{4.118}$$

By suitably stretching or compressing the time scale, which has no physical significance, we can make the constant in (4.118) equal to c; then

$$\dot{t} = \frac{1}{c} \tag{4.119}$$

To display the centrifugal force we consider a momentary radial motion in the rotating system; that is, we set $\dot{\varphi} = 0$ and insert (4.119) in Eq. (4.111) to get

$$\ddot{r} = \frac{rw^2}{c^2} \tag{4.120}$$

Using the proper-time interval of the particle in the barred system $d\tau = ds/c$, we then have

$$\frac{d^2r}{d\tau^2} = rw^2 \tag{4.121}$$

which we recognize as the familiar classical expression for centrifugal acceleration, except that the proper-time interval $d\tau$ replaces the coordinate-time interval dt.

Similarly, we can obtain the Coriolis acceleration by differentiating (4.114) with respect to s and taking $\dot{\varphi}$ to be instantaneously zero:

$$\begin{gathered} 2\dot{r}\dot{\varphi} + r\ddot{\varphi} + \frac{2w\dot{r}}{c} = 0 \\ r\ddot{\varphi} + 2\,\frac{w\dot{r}}{c} = 0 \end{gathered} \tag{4.122}$$

In terms of the proper-time interval $d\tau$ this may be written as

$$r\,\frac{d^2\varphi}{d\tau^2} + 2w\,\frac{dr}{d\tau} = 0 \tag{4.123}$$

which we also recognize as the classical Coriolis acceleration, again with the proper-time interval $d\tau$ replacing the coordinate-time interval dt.

From the above results it appears reasonable at this point to adopt tentatively the geodesic equations as the equations of motion in a four-

dimensional Riemann space—at least in the absence of forces. This has indeed proved to agree with classical theory for the example of the inertial system which we considered in the preceding paragraphs; furthermore, we obtained a reasonable generalization of the familiar Coriolis and centrifugal accelerations when we applied the geodesic equations to the example of the rotating coordinate system. In addition, the variational problem $\delta\int ds = 0$ is an invariant expression and provides an elegant and invariant characterization of the path of a particle. In the next section we shall investigate further the link between geometry and physics provided by the geodesic equations by an application to a more general metric space than the Lorentz form we considered in this section. This investigation will provide further justification for the use of the geodesic equations of motion.

4.3 Gravity as a Metric Phenomenon

In the last section we found that using the Lorentz metric and the geodesic equation as equations of motion, we obtained, in the examples considered, a reasonable description of the force-free motion of a particle in a four-dimensional Riemann space. In this section we shall attempt to show by an approximation procedure that the effect of a gravitational field of force can be described by again using the geodesic equation of motion and by allowing the metric tensor to differ "somewhat" from the Lorentz metric. The success of this procedure will serve as justification for linking the physical force of gravity with the non-Lorentzian nature of space and for using the geodesic equations as equations of motion for particles in a gravitational field.

Since the geometry of the real world is Euclidean so far as any ordinary physical measurement is concerned, the space-time metric must be very close indeed to Lorentzian. However, it is just the very minute departure from the Lorentzian metric which we desire to show is the agent of gravitational effects. Thus we shall consider a *time-independent* metric tensor of the form

$$g_{\mu\nu} = \eta_{\mu\nu} + \epsilon\gamma_{\mu\nu} \qquad \epsilon = \text{small constant} \tag{4.124}$$

where $\eta_{\mu\nu}$ is the Lorentz metric tensor and $\epsilon\gamma_{\mu\nu}$ represents a very small time-independent perturbation which is due to the presence of a gravitating body and goes to zero very far from the body. We term this $g_{\mu\nu}$ a nearly Lorentzian metric tensor.

To show that the $\epsilon\gamma_{\mu\nu}$ term is indeed the agent of gravitational forces, we shall apply the geodesic equations of motion to a Riemann space with the above metric. Furthermore, to make a close connection with classical theory, we shall suppose that the velocity of the particle along the geodesic is much less than c, or equivalently, that $\beta = v/c$ is very small; then in our approximate calculations we shall retain only first-order terms in ϵ and β, dropping terms of order ϵ^2, β^2, $\epsilon\beta$, and higher.

Using the coordinates of special relativity and the nearly Lorentzian metric tensor (4.124), we immediately obtain the line element

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 + \epsilon\gamma_{\mu\nu}\, dx^\mu\, dx^\nu \tag{4.125}$$

Thus

$$\left(\frac{ds}{dt}\right)^2 = c^2 - v^2 + \epsilon\gamma_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} = c^2\left(1 - \beta^2 + \epsilon\gamma_{\mu\nu}\frac{dx^\mu}{dx^0}\frac{dx^\nu}{dx^0}\right) \tag{4.126}$$

To first order in ϵ and β this is

$$\left(\frac{ds}{dt}\right)^2 \cong c^2(1 + \epsilon\gamma_{00}) \tag{4.127}$$

We next apply the same approximation to the differential equations of a geodesic:

$$\frac{d^2x^\alpha}{ds^2} + \begin{Bmatrix}\alpha\\ \eta\ \tau\end{Bmatrix}\frac{dx^\eta}{ds}\frac{dx^\tau}{ds} = 0 \tag{4.128}$$

Consider the second term on the left side of this equation. Since $\eta_{\mu\nu}$ is constant in space-time, it is evident from the form of the metric (4.125) that each Christoffel symbol contains a factor ϵ. Using the expression (4.127) for $(ds/dt)^2$, we can write

$$\frac{dx^\eta}{ds}\frac{dx^\tau}{ds} = \frac{dx^\eta}{dt}\frac{dx^\tau}{dt}\left(\frac{dt}{ds}\right)^2 = \frac{dx^\eta}{dt}\frac{dx^\tau}{dt}\frac{1}{c^2(1 + \epsilon\gamma_{00})} \tag{4.129}$$

If neither η nor τ is zero, this can be written in terms of the τ and η components of the velocity v as

$$\frac{dx^\eta}{dt}\frac{dx^\tau}{dt}\frac{1}{c^2(1 + \epsilon\gamma_{00})} = \frac{v^\eta}{c}\frac{v^\tau}{c}\frac{1}{(1 + \epsilon\gamma_{00})} \tag{4.130}$$

which is of order β^2. If only one of the indices η or τ is not zero, the expression is clearly of order β. Thus, unless $\eta = \tau = 0$, the product $\begin{Bmatrix}\alpha\\ \eta\ \tau\end{Bmatrix}\frac{dx^\eta}{ds}\frac{dx^\tau}{ds}$ is of order $\epsilon\beta$ or higher, and is to be neglected in our approximation scheme. The geodesic equations to first order in ϵ and β are then

$$\frac{d^2x^\alpha}{ds^2} + \begin{Bmatrix}\alpha\\ 0\ 0\end{Bmatrix}\left(\frac{dx^0}{ds}\right)^2 = 0 \tag{4.131}$$

By virtue of (4.127) these may also be written within our degree of approximation as

$$\frac{d^2x^\alpha/dt^2}{c^2(1 + \epsilon\gamma_{00})} + \frac{\begin{Bmatrix}\alpha\\ 0\ 0\end{Bmatrix}}{1 + \epsilon\gamma_{00}} = 0 \tag{4.132}$$

since all terms neglected in the transition from (4.131) to (4.132) contain factors $\epsilon\beta$. Equivalently,

$$\frac{d^2x^\alpha}{dt^2} + \begin{Bmatrix}\alpha\\ 0\ 0\end{Bmatrix}c^2 = 0 \tag{4.133}$$

We do not need the differential equation for $x^0 = ct$, and indeed we shall have to test the compatibility $\begin{Bmatrix}0\\ 0\ 0\end{Bmatrix} = 0$.

In order to simplify this approximate equation further, we shall calculate the Christoffel symbol $\begin{Bmatrix}\alpha\\ 0\ 0\end{Bmatrix}$ explicitly. By definition the corresponding Christoffel symbol of the first kind is

$$[00, \lambda] = \tfrac{1}{2}(g_{0\lambda|0} + g_{\lambda 0|0} - g_{00|\lambda}) \tag{4.134}$$

Since, by our assumption of a time-independent metric, $g_{\mu\nu}$ is independent of x^0 and $\eta_{\mu\nu}$ is a constant in all the variables x^μ, this becomes

$$[00, \lambda] = -\tfrac{1}{2}g_{00|\lambda} = -\tfrac{1}{2}\epsilon\gamma_{00|\lambda} \tag{4.135}$$

Raising the index λ to obtain the Christoffel symbol of the second kind and ignoring terms of order ϵ^2, we then obtain

$$\begin{Bmatrix}\alpha\\ 0\ 0\end{Bmatrix} = g^{\alpha\lambda}[00, \lambda] = -\tfrac{1}{2}g^{\alpha\lambda}\epsilon\gamma_{00|\lambda} = -\tfrac{1}{2}g^{(L)\alpha\lambda}\epsilon\gamma_{00|\lambda} \tag{4.136}$$

Separating time and space components, we obtain for $\alpha = 0$

$$\begin{Bmatrix}0\\ 0\ 0\end{Bmatrix} = -\tfrac{1}{2}\epsilon\gamma_{00|0} = 0 \tag{4.137}$$

because of our assumption of time independence. This is consistent with the definition $x^0 = ct$ and the validity of (4.133) for $\alpha = 0$; we recognize that our two physical assumptions, (1) time independence of the metric and (2) smallness of velocities in our coordinate system, are interdependent. This is the reason for the compatibility of (4.137) and (4.133). For $\alpha = i = 1, 2, 3$, we have

$$\left\{ {i \atop 0\ 0} \right\} = \tfrac{1}{2}\epsilon\gamma_{00|i} \tag{4.138}$$

Using this approximation for the Christoffel symbol, we can write the geodesic equations (4.133) as

$$\frac{d^2x^i}{dt^2} = -\frac{c^2}{2}\epsilon\gamma_{00|i} \tag{4.139}$$

In three-dimensional vector notation this is

$$\frac{d^2\mathbf{x}}{dt^2} = -\frac{c^2}{2}\epsilon\nabla\gamma_{00} \tag{4.140}$$

This is simply Newton's equation of motion in a classical gravitational field derived from a scalar potential if we identify the scalar potential as

$$\varphi = \frac{c^2}{2}\epsilon\gamma_{00} \tag{4.141}$$

Conversely, given the classical potential φ, the motion of a particle will be along a four-dimensional geodesic if the g_{00} term of the metric tensor has the form

$$g_{00} = 1 + \frac{2\varphi}{c^2} \tag{4.142}$$

The other components do not enter in our approximation, except through the assumption that they are time-independent and nearly Lorentzian.

Let us summarize the preceding results: If we ignore second-order terms in ϵ and β (the weak field and low velocity limit), then the geodesic equation (a purely geometric relation) is equivalent to Newton's equation (4.140) (a purely mechanistic relation), provided the g_{00} of the metric tensor satisfies the relation (4.142). This equivalence provides rather good justification for the use of a non-Lorentzian metric to describe a gravitational field and for the use of the geodesic equations of motion in

the resulting Riemann space. Furthermore, the approximate relation (4.142) for the g_{00} term of the metric tensor will itself be useful for several reasons: (1) It will allow us to relate certain constants which appear formally in the later developments of the relativity theory to familiar classical quantities, such as mass and the gravitational constant. (2) We may check, when we solve the gravitational field equations of Einstein in Chap. 8, that the g_{00} component is consistent with our approximate result. (3) We can predict the red shift of spectral lines in a gravitational field, using only the above results in the framework of special relativity. This will be the subject of the next section.

4.4 The Red Shift

Without going beyond the preliminary notions we have developed in this chapter, we can predict a rather interesting effect of the gravitational field: the slowing down of time in the field and the consequent red shift of spectral lines emitted by atoms located on massive bodies. The effect has been tested by experiment and been rather well verified; we thus have experimental justification for the basic theoretical concepts we have set forth in this chapter.

Consider, for example, a light wave emitted on the sun and received on the earth. Let the gravitational potential at the surface of the sun be φ_s. Using (4.94) and the approximate g_{00} given in Eq. (4.142), proper-time intervals are related to coordinate-time intervals by the equation

$$d\tau_s = \sqrt{g_{00}(x_s^\mu)}\,dt = \left(1 + \frac{2\varphi_s}{c^2}\right)^{1/2} dt \tag{4.143}$$

Similarly, on the earth, proper-time intervals are related to coordinate-time intervals by

$$d\tau_e = \sqrt{g_{00}(x_e^\mu)}\,dt = \left(1 + \frac{2\varphi_e}{c^2}\right)^{1/2} dt \tag{4.144}$$

where φ_e is the value of the gravitational potential on the earth. Suppose now n waves of frequency ν_0 are emitted in proper time $\Delta\tau_s$ from an atom on the sun. Then

$$n = \nu_0\,\Delta\tau_s \tag{4.145}$$

On the earth one certainly receives n waves, but the frequency and time duration of the wave train have changed. Using a frequency-duration relation for the earth analogous to (4.145) for the sun,

$$n = \nu_e \,\Delta\tau_e \tag{4.146}$$

we obtain, since n is a constant,

$$\nu_0 \,\Delta\tau_s = \nu_e \,\Delta\tau_e \tag{4.147}$$

Thus

$$\nu_e = \nu_0 \frac{\Delta\tau_s}{\Delta\tau_e} \tag{4.148}$$

From (4.143) the coordinate-time duration of the wave corresponding to $\Delta\tau_s$ is

$$\Delta t = \frac{\Delta\tau_s}{\sqrt{g_{00}(x_s^\mu)}} = \frac{\Delta\tau_s}{\sqrt{1 + 2\varphi_s/c^2}} \tag{4.149}$$

We suppose that the coordinate-time duration of the wave Δt is the *same* on the earth as on the sun. [See Sec. 4.2 on proper-time and coordinate-time, especially (4.95).] Equation (4.144) then gives

$$\Delta t = \frac{\Delta\tau_e}{\sqrt{g_{00}(x_e^\mu)}} = \frac{\Delta\tau_e}{\sqrt{1 + 2\varphi_e/c^2}} \tag{4.150}$$

By virtue of this and (4.149), we then have

$$\frac{\Delta\tau_s}{\Delta\tau_e} = \frac{\sqrt{g_{00}(x_s^\mu)}}{\sqrt{g_{00}(x_e^\mu)}} = \left(\frac{1 + 2\varphi_s/c^2}{1 + 2\varphi_e/c^2}\right)^{1/2} \tag{4.151}$$

Substitution of this into (4.148) gives

$$\nu_e = \nu_0 \frac{\sqrt{g_{00}(x_s^\mu)}}{\sqrt{g_{00}(x_e^\mu)}} = \nu_0 \left(\frac{1 + 2\varphi_s/c^2}{1 + 2\varphi_e/c^2}\right)^{1/2} \tag{4.152}$$

Expanding to first order in the small quantities φ_s/c^2 and φ_e/c^2, we obtain

$$\frac{\nu_e - \nu_0}{\nu_0} = \frac{\varphi_s - \varphi_e}{c^2} \tag{4.153}$$

or in briefer notation,

$$\frac{\Delta\nu}{\nu_0} = \frac{\Delta\varphi}{c^2} \tag{4.154}$$

Since the sun is at a large negative potential relative to the earth, we see that $\Delta\varphi$ is negative. Thus the frequency of light *decreases* as it leaves the sun, and when it is received on earth, we see a shift toward the red end of the spectrum. It is as if the atoms of the sun were vibrating in slow motion when we view them from the earth. Of course, there is nothing special about using the earth and sun as the two points considered, and we can just as well use two points at different heights on the earth if our measurement is precise enough to detect the correspondingly small shift.

Alternative derivations of the red shift formula. Using the equivalence principle in a very direct way, Einstein made the first derivation of the red shift by considering a system rotating at constant angular velocity w such as we used in Sec. 4.2. The linear velocity of a fixed point at radius r is then $v = wr$, and the centrifugal potential is

$$\varphi_{\text{cent}} = -\tfrac{1}{2}w^2r^2 \tag{4.155}$$

As obtained in Sec. 4.2, the relation between time intervals on the rotating system and external inertial intervals is, in agreement with special relativity,

$$d\tau = \left(1 - \frac{v^2}{c^2}\right)^{1/2} dt = \left(1 - \frac{w^2r^2}{c^2}\right)^{1/2} dt = \left(1 + \frac{2\varphi_{\text{cent}}}{c^2}\right)^{1/2} dt \tag{4.156}$$

But by the principle of equivalence, the centrifugal potential should be experimentally indistinguishable in the small from a gravitational potential, and the above relation should hold equally well for a gravitational potential φ. Proceeding as before, we then obtain the red shift formula from the above, since (4.156) and (4.143) are formally identical.

For the case of the terrestrial red shift, where the points under consideration are close together, another derivation is possible. Let us replace the gravitational field by an acceleration of the coordinate system. That is, instead of placing our apparatus in the earth's field with potential gradient $g = 980$ cm/sec², we put it in an elevator or rocket accelerating at g in free space, which (according to the equivalence principle) will give identical physical results (Fig. 4.1). The time it takes a light wave to travel between sender and receiver is roughly d/c. But in this time the receiver has increased its velocity over that of the sender when the wave was emitted by an amount $(d/c)g$. Thus there is a consequent Doppler shift corresponding to $\Delta v/c = (d/c^2)g$. The frequency shift $\Delta\nu/\nu = -\Delta v/c$ is then

$$\frac{\Delta\nu}{\nu} = -\frac{gd}{c^2} \tag{4.157}$$

If we note that $-gd$ is the change in the gravitational potential $\Delta\varphi$ over a distance d, we can write as before $\Delta\nu/\nu = \Delta\varphi/c^2$.

It is important to note that the two alternative derivations of the red shift which have been presented here rely on the validity of the *principle of equivalence*, but not on the field equations of the general theory of relativity; the derivation of the red shift does not require the use of Einstein's equations, which will be introduced in the next chapter.

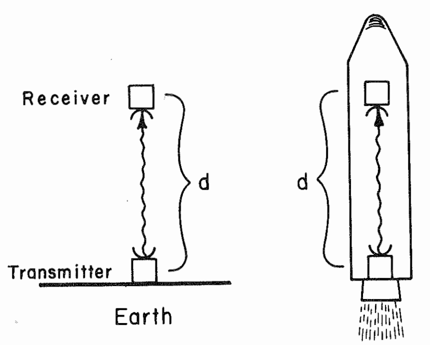
Yet a third derivation of the red shift is possible which does not rely explicitly on the principle of equivalence but only on the mass-energy relation $E = mc^2$ and the interpretation of light as energetic quanta with an effective mass m. According to the mass-energy relation, a light quantum of energy $E = h\nu$ will have effective mass

$$m = \frac{E}{c^2} = \frac{h\nu}{c^2} \qquad h = \text{Planck's constant} \tag{4.158}$$

Accordingly, the sum of potential and kinetic energy of the quantum at a point where the gravitational potential is φ will be $h\nu + m\varphi$. Thus, when the quantum moves through a potential difference $\Delta\varphi$, say, between the sun (s) and the earth (e), we have the energy-balance equation

$$h\nu_e + m\varphi_e = h\nu_0 + m\varphi_s \tag{4.159}$$

Fig. 4.1



Using the effective mass relation (4.158), we have

$$h\nu_e - h\nu_0 = m(\varphi_s - \varphi_e) = \frac{h\nu_0}{c^2}(\varphi_s - \varphi_e) \tag{4.160}$$

which gives, upon canceling the h, the previous result (4.154):

$$\frac{\Delta\nu}{\nu_0} = \frac{\Delta\varphi}{c^2} \tag{4.161}$$

Experimental tests of the red shift. Historically, the initial interest in the red shift equation (4.154) centered about its experimental verification through spectral measurements of the sun and other stars; this would be a rather direct verification of the principle of equivalence. Such astronomical measurements are, however, difficult to make, and their interpretation is apt to be ambiguous, because of the presence of nongravitational effects on a stellar surface such as Doppler shifts in the high-temperature gas, intense electromagnetic radiation fields, possible high electric fields due to gas ionization, vertical currents in the stellar gas, high gas pressure, etc. As a result it has been difficult to isolate accurately that part of the observed spectral shift which is attributable to the gravitational red shift. We shall mention only one recent result. Brault and Dicke (Brault, 1962; Dicke, 1964) used direct electronic techniques without appealing to the usual photographic-plate techniques of astronomers. They measured the displacement of the center of the broad D_1 line of sodium as a function of the radial distance across the solar disk. They found to within 5 per cent accuracy that the shift was constant (after corrections for the small line asymmetry) and agreed with the theoretical value.

Since it has been impossible to place complete confidence in astronomical measurements, there has naturally been considerable interest in the possibility of a terrestrial test of the red shift. This is a very difficult task, for the expected shift over a vertical distance of, say, 100 ft, is only of the order of 10^{-15}. Fortunately, the discovery of the Mössbauer effect in 1958 (Mössbauer, 1958a and b, 1959) gave a method of producing and detecting gamma rays which are monochromatic to 1 part in 10^{12} and made a terrestrial test feasible.

In general, it is difficult to measure very small shifts in the gamma-ray spectral lines of nuclei; this is because a nucleus in a crystal is usually able to recoil when emitting a gamma ray, and one therefore observes a Doppler effect in the observed spectra. Mössbauer found that in some crystals such as Fe^{57} the whole crystal instead of one nucleus can take up the recoil momentum. This makes the emission effectively recoilless, and one can observe very narrow spectral lines. Indeed, in Fe^{57}, there is a line near 14.4 kev with a fractional width of 10^{-12}. If a radioactive emitting sample of Fe^{57} is placed near a thin nonradioactive absorbing sheet of the same material, one then finds that a large fraction of the gamma radiation falling on the sheet is *resonantly absorbed*. If, however, either the absorber or emitter is moved at a velocity of only a few millimeters per second, the consequent Doppler shift becomes as large as the line width and the resonant absorption falls off very rapidly. This gives one a very accurate tool for the detection of quite small frequency shifts.

In the red shift experiment (Pound and Rebka, 1960; and Pound and Snider, 1964) emitter and absorber were placed at opposite ends of a vertical 72-ft tower. Gamma rays emitted at the bottom then suffered a gravitational red shift in traveling to the absorber at the top and, as a result, were less favorably absorbed. By moving the emitter upward at a small velocity, a compensating Doppler shift was produced which restored resonant absorption. A measurement of the emitter velocity then allowed a calculation of the ratio $\Delta\nu/\nu$. The experimental result obtained is 0.997 ± 0.008 times the predicted shift of 4.92×10^{-15}. This result represents verification of the correctness of the red shift equation (4.154) to better than 1 per cent.

This ends for the present the consideration of how the gravitational field, in the form of a non-Lorentzian metric, influences the matter (or light) in its vicinity. In the next chapter we shall consider the converse problem, how the matter influences the metric structure of space, and begin the discussion of the gravitational field equations in free space.

Exercises

4.1 Show that in Lorentz space

$$F^{\sigma\tau}F_{\sigma\tau} = 2(\mathbf{H}^2 - \mathbf{E}^2) \qquad {}^*F^{\alpha\beta}F_{\alpha\beta} = 4\,\mathbf{E}\cdot\mathbf{H}$$

Are there any more bilinear invariants that can be constructed from the components of the electromagnetic field? What are the values of the above invariants for a plane wave?

4.2 Define the following complex tensor from the Maxwell tensor and its dual tensor

$$\omega_{\mu\nu} = F_{\mu\nu} + i({}^*F_{\mu\nu})$$

Express Maxwell's equations in terms of $\omega_{\mu\nu}$.

4.3 The tensor introduced above has the remarkable property that the bilinear invariants of electromagnetic theory discussed in Exercise 4.1 are simply accommodated in a single complex invariant. To see this show that

$$\omega^{\alpha\beta}\omega_{\alpha\beta} = 2(F_{\sigma\tau}F^{\sigma\tau}) + i({}^*F^{\alpha\beta}F_{\alpha\beta})$$

$$\omega^{\alpha\beta}\bar{\omega}_{\alpha\beta} = 0$$

where $\bar{\omega}_{\alpha\beta}$ is the complex conjugate of $\omega_{\alpha\beta}$. What is $\bar{\omega}^{\alpha\beta}\bar{\omega}_{\alpha\beta}$?

4.4 Consider Maxwell's equations in special relativity. Show that by a gauge transformation one may assure that

$$A^{\mu}{}_{|\mu} = 0$$

This is termed the Lorentz gauge condition. Using such a gauge, show that A^μ satisfies the equation

$$\Box^2 A^\mu \equiv A^{\mu}{}_{|\alpha|\beta}g^{\alpha\beta} = s^\mu$$

4.5 In free space show that $F^{\mu\nu}$ satisfies the wave equation whereas A^μ satisfies it only if the Lorentz condition is imposed. (Work in Lorentz space: is the exercise true in an arbitrary Riemann space?)

4.6 The coordinate transformation considered in special relativity is linear. What is it explicitly for motion in the x direction? What are the transformation laws for contravariant and covariant special relativistic four-vectors? What are they for second-rank tensors? Obtain the transformation laws for the electromagnetic fields under the Lorentz transformation. State them in the form

$$E'_{\parallel} = F(E,B) \qquad H'_{\parallel} = K(E,B)$$

$$E'_{\perp} = G(E,B) \qquad H'_{\perp} = J(E,B)$$

where $\parallel$ means parallel to the direction of relative motion and $\perp$ means perpendicular to the direction of relative motion.

4.7 In a weak gravitational field represented by a first-order metric with $g_{00} = 1 - 2\varphi/c^2$ show that the Lorentz time-dilation factor should be replaced by $(1 - v^2/c^2 - 2\varphi/c^2)^{-1/2}$.

Problems

A useful reference for these problems is Vishveshwara, 1968.

4.1 Let k^α represent a field of null geodesics; i.e., the lines with tangent vectors $k^\alpha = dx^\alpha/dq$ are geodesics, and $k^\mu k_\mu = 0$. Show that this implies that $k^\alpha{}_{\|\beta}k^\beta = 0$ if q is one of the distinguished parameters discussed in Sec. 2.3. Show moreover that by a suitable normalization k^α can be interpreted as the wave vector of a photon. See Sec. 6.5.

4.2 A stationary metric has, by definition, a timelike Killing vector ξ^μ, which can be normalized to yield a unit vector $u^\alpha = \xi^\alpha/(\xi^\mu\xi_\mu)^{1/2}$. Show that u^α may be interpreted as the four-velocity of an observer at rest. Show that such an observer will observe the photons discussed in Prob. 4.1 to have frequencies $u^\alpha k_\alpha$.

4.3 Along a null geodesic line show that $(d/dq)\ (k_\alpha\xi^\alpha) = 0$ and that the red shift of a photon emitted at s and observed at o by stationary observers may be written

$$\frac{\nu_o}{\nu_s} = \frac{(k_\alpha u^\alpha)_o}{(k_\alpha u^\alpha)_s} = \frac{(\xi^\mu\xi_\mu)_s{}^{1/2}}{(\xi^\beta\xi_\beta)_o{}^{1/2}}$$

This covariant statement is equivalent to (4.152).

Bibliography

Textbooks on special relativity:

Aharoni, J. (1959): "The Special Theory of Relativity," Oxford.
Bergmann, P. G. (1942): "An Introduction to the Theory of Relativity," New York.
Bergmann, P. G. (1962): The Special Theory of Relativity, in Encyclopedia of Physics, vol. 4, Berlin-Göttingen-Heidelberg, pp. 108–202.
Laue, M. von (1955): "Die Relativitätstheorie," vol. 1, "Spezielle Relativitätstheorie," Brunswick.
Papapetrou, A. (1955): "Spezielle Relativitätstheorie," Berlin.
Sommerfeld, A. (1952): "Lectures on Theoretical Physics," vol. 3, "Electrodynamics," New York.
Synge, J. L. (1956): "Relativity: The Special Theory," Amsterdam.

On characteristics:

Courant, R., and D. Hilbert (1962): "Methods of Mathematical Physics," vol. 2, "Partial Differential Equations," New York.
Fock, V. (1959): "The Theory of Space, Time and Gravitation," New York.

Experimental tests of red shift:

Adam, M. G. (1955): Interferometric Measurement of Wave Lengths, II, *Monthly Notices Roy. Astron. Soc.*, **115**:405–421.
Blamont, J. E., and F. Roddier (1961): Precise Observation of the Profile of the Strontium Resonance Line: Evidence for the Gravitational Red Shift on the Sun, *Phys. Rev. Letters*, **7**:437–439.
Brault, J. (1962): Thesis, Princeton University, Princeton, N.J.
Cranshaw, T. E., J. P. Schiffer, and A. B. Whitehead (1960): Measurement of the Gravitational Red Shift Using the Mössbauer Effect in Fe^{57}, *Phys. Rev. Letters*, **4**:163–164.
Dicke, R. H. (1964): Experimental Relativity, in B. Dewitt (ed.), "Relativity, Groups and Topology," New York.
Finlay-Freundlich, E. (1954): Red Shifts in the Spectra of Celestial Bodies, *Philos. Mag.*, **45**:303–319.
Frauenfelder, H. (1962): "The Mössbauer Effect," New York.
Hay, H. J., J. P. Schiffer, T. E. Cranshaw, and P. A. Egelstaff (1960): Measurement of the Red Shift in an Accelerated System Using the Mössbauer Effect in Fe^{57}, *Phys. Rev. Letters*, **4**:165–166.
Mössbauer, R. L. (1958a): Kernresonanzfluoreszenz von Gammastrahlung in Ir^{191}, *Z. Physik*, **151**:124–143.
Mössbauer, R. L. (1958b): Kernresonanzabsorption von Gammastrahlung in Ir^{191}, *Naturwissenschaften*, **45**:538–539.
Mössbauer, R. L. (1959): Kernresonanzabsorption von γ-Strahlung in Ir^{191}, *Z. Naturforsch.*, **14a**:211–216.
Pound, R. V., and G. A. Rebka (1959): Gravitational Red Shift in Nuclear Resonance, *Phys. Rev. Letters*, **3**:439–441.
Pound, R. V., and G. A. Rebka (1960): Apparent Weight of Photons, *Phys. Rev. Letters*, **4**:337–341.
Pound, R. V., and J. L. Snider (1964): Effect of Gravity on Nuclear Resonance, *Phys. Rev. Letters*, **13**:539–540.
Schröter, E. H. (1956): Deutung der Rotverschiebung . . . der Sonnenatmosphäre, *Z. Astrophys.*, **41**:141–181.
Vishveshwara, C. V. (1968): Generalization of the "Schwarzschild Surface" to Arbitrary Static and Stationary Metrics, *J. Math. Phys.*, **9**:1319.

CHAPTER

5

The Gravitational Field Equations in Free Space

From our investigations in the preceding chapter we can state the following two conclusions: (1) In the absence of forces, mechanical phenomena can be suitably described in a differential geometric framework by using a Lorentz metric and a geodesic equation of motion. (2) The effect of gravitational forces can be included in the differential geometric framework by using a non-Lorentzian metric and retaining the geodesic equation of motion; the metric in this case differs only slightly from the Lorentz metric in regions where the gravitational potential is nonzero, and is Lorentzian where the gravitational potential is zero. The program of this chapter is to obtain field equations for the metric tensor in matter-free regions of space; that is, we wish to answer the question of how matter affects the metric structure of the free space in its vicinity.

5.1 Criteria for the Field Equations

Our first criterion, as could be expected from our entire approach, is that the field equations (and the whole theory!) be phrased in covariant tensor form. This is Einstein's well-known *principle of covariance*, and is seen to be desirable for the following reasons:

1. The principle of equivalence implies that accelerated systems must be considered to be quite as respectable as inertial systems, so we demand that physical laws do not distinguish between the two. This will clearly be so if the laws are tensor laws, for then the system of coordinates does not enter the equations at all.

2. From Chap. 4 we know that, in a four-dimensional tensor formulation, both fictitious forces and gravitational forces appear as Christoffel symbols in the geodesic equations of motion. The fact that they appear mathematically alike is a very desirable characteristic of the tensor approach, for according to the equivalence principle, gravitational and fictitious (inertial) forces are indistinguishable "in the small."

3. It was pointed out first by Kretschmann (Kretschmann, 1917) that on purely mathematical grounds any (tentative) physical law expressed in a particular coordinate system can be brought into a covariant tensor form. However, in physics, one wants to keep the number of quantities (observables) entering the equations to a minimum; this will single out certain laws for which no new formal tensor quantities will have to be introduced in the process of making the formulas covariant. The principle of covariance thereby provides a purely formal but successful guide to choosing physical laws on the basis of elegance and simplicity. Applying it to the problem of gravity gives the first criterion: (*a*) *The gravitational equations should be phrased in covariant tensor form.*

The classical theory of gravitation in free space can be basically stated in two equations, namely, Newton's second law in a gravitational field with a potential φ,

$$\frac{d^2x^i}{dt^2} = -\frac{\partial\varphi}{\partial x^i} \tag{5.1}$$

and Laplace's equation for the potential:

$$\sum_{i=1}^{3} \frac{\partial^2\varphi}{\partial x^{i2}} = 0 \tag{5.2}$$

(The Latin indices run from 1 to 3 by our standing convention.) Equation (5.1) tells how matter moves in a given gravitational field. Equation (5.2) describes how matter (or rather its absence) determines the gravitational field; indeed, (5.2) is a special case of the Poisson-Laplace equation $\nabla^2\varphi = -4\pi\rho$. We saw in Sec. 4.3 that the geodesic equation of motion which we desire to use in the relativity theory reduces to the classical law (5.1) *if* we identify $g_{00} \cong 1 + 2\varphi/c^2$. This provides us with a second clue to the nature of the relativity field equations; indeed, since Laplace's equation involves the second derivatives of φ, we might expect that the general relativistic field equations will involve the second derivatives of g_{00}. Because of the tensor form of the equations, this implies that the second derivatives of all the components of $g_{\mu\nu}$ should appear. Our inclination to look for second-order equations is strengthened also by the fact that most of the differential field equations of classical physics are of second order. Thus our second criterion is: (*b*) *The field equations should be of second order in the components of the metric tensor.*

We found in Chap. 4 that, in the absence of forces, we can describe mechanics using a Lorentz metric, so we adopt as a third criterion: (*c*) *For the case where all space is empty of matter* (and there is consequently no gravitational field), *the field equation must admit the Lorentz metric as a particular solution.* This last criterion will be somewhat modified in Chap. 10, when we consider large-scale cosmological problems.

Our fourth criterion is intended to guarantee a unique solution of the field equations. If we have a differential equation of the form

$$F(y^{(n)}, \ldots, y, x) = 0$$

which can be solved for $y^{(n)}$, it is clear that the differential equation can have a unique solution only if $y^{(n)}$ is uniquely determined by $y^{(n-1)}$, $y^{(n-2)}$, etc. This will surely be the case if $y^{(n)}$ enters linearly; indeed, if the function F is algebraic, linearity in $y^{(n)}$ is a necessary and sufficient condition for uniqueness of a solution. Thus, for the field equations, we ask that the second derivatives of $g_{\mu\nu}$ enter linearly; that is, the equations should be quasi-linear. The final criterion can then be stated: (*d*) *The field equations should be quasi-linear.*

Other sets of guiding criteria for developing the field equations are of course possible, but the set of four criteria we have stated above are particularly simple, and will prove to be an adequate guide in the development of the following sections.

5.2 The Riemann Curvature Tensor

The first guide we shall invoke is criterion (*c*). Since the Lorentz metric must be one very important solution of the gravitational field equations, we shall try to find a necessary condition, expressed in the form of a differential field equation, for a space to have a Lorentz metric.

Let us note first an important property of the Lorentz metric: in the coordinates of special relativity its components are *constant over all space.* This property will be essential in providing our criteria for a space to have a Lorentz metric. If a space has a Lorentz metric, then by assumption there exists a system where the metric is constant and has the Lorentz form. That is, there exists a geodesic system *in the large.* All the Christoffel symbols are *everywhere* zero in that system; consequently the ordinary derivatives and the covariant derivatives of any vector ξ^α are equal:

$$\xi^\alpha{}_{\|\beta} = \xi^\alpha{}_{|\beta} \tag{5.3}$$

Indeed, it is evident that covariant and ordinary derivatives of *all* orders are equal if and only if the Christoffel symbols are *everywhere* zero, so we also have

$$\xi^{\alpha}{}_{\|\beta\|\gamma} = \xi^{\alpha}{}_{|\beta|\gamma} \tag{5.4}$$

and similarly for derivatives of all orders. Since, for any vector field, the order of ordinary differentiation is irrelevant, it follows that the order of covariant differentiation is also irrelevant. That is,

$$\xi^{\alpha}{}_{\|\beta\|\gamma} - \xi^{\alpha}{}_{\|\gamma\|\beta} = 0 \tag{5.5}$$

But this is a tensor equation; since it is true in one system, it must be true in *all systems*, not just the geodesic system. Thus we have the rule that *when* a space admits a Lorentz metric, (5.5) holds. The differential equation (5.5) is consequently a necessary condition for a space to admit a Lorentz metric. A little manipulation will put this equation in more useful form. For convenience, let us define a tensor $t^{\alpha}{}_{\beta}$:

$$t^{\alpha}{}_{\beta} = \xi^{\alpha}{}_{\|\beta} = \xi^{\alpha}{}_{|\beta} + \begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix} \xi^{\eta} \tag{5.6}$$

Then

$$t^{\alpha}{}_{\beta\|\gamma} = \xi^{\alpha}{}_{\|\beta\|\gamma} = t^{\alpha}{}_{\beta|\gamma} + \begin{Bmatrix} \alpha \\ \tau\ \gamma \end{Bmatrix} t^{\tau}{}_{\beta} - \begin{Bmatrix} \lambda \\ \beta\ \gamma \end{Bmatrix} t^{\alpha}{}_{\lambda} \tag{5.6'}$$

Inserting (5.6) into (5.6′), we obtain

$$\begin{aligned} \xi^{\alpha}{}_{\|\beta\|\gamma} = \xi^{\alpha}{}_{|\beta|\gamma} &+ \begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix}_{|\gamma} \xi^{\eta} + \begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix} \xi^{\eta}{}_{|\gamma} + \begin{Bmatrix} \alpha \\ \tau\ \gamma \end{Bmatrix} \xi^{\tau}{}_{|\beta} \\ &+ \begin{Bmatrix} \alpha \\ \tau\ \gamma \end{Bmatrix} \begin{Bmatrix} \tau \\ \beta\ \eta \end{Bmatrix} \xi^{\eta} - \begin{Bmatrix} \lambda \\ \beta\ \gamma \end{Bmatrix} t^{\alpha}{}_{\lambda} \end{aligned} \tag{5.7}$$

Interchanging β and γ, we obtain a similar expression.

$$\begin{aligned} \xi^{\alpha}{}_{\|\gamma\|\beta} = \xi^{\alpha}{}_{|\gamma|\beta} &+ \begin{Bmatrix} \alpha \\ \gamma\ \eta \end{Bmatrix}_{|\beta} \xi^{\eta} + \begin{Bmatrix} \alpha \\ \gamma\ \eta \end{Bmatrix} \xi^{\eta}{}_{|\beta} + \begin{Bmatrix} \alpha \\ \eta\ \beta \end{Bmatrix} \xi^{\eta}{}_{|\gamma} \\ &+ \begin{Bmatrix} \alpha \\ \tau\ \beta \end{Bmatrix} \begin{Bmatrix} \tau \\ \gamma\ \eta \end{Bmatrix} \xi^{\eta} - \begin{Bmatrix} \lambda \\ \gamma\ \beta \end{Bmatrix} t^{\alpha}{}_{\lambda} \end{aligned} \tag{5.8}$$

The difference of the two expressions then gives the left side of (5.5) in the form

$$\begin{aligned} \xi^{\alpha}{}_{\|\beta\|\gamma} - \xi^{\alpha}{}_{\|\gamma\|\beta} &= \begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix}_{|\gamma} \xi^{\eta} - \begin{Bmatrix} \alpha \\ \gamma\ \eta \end{Bmatrix}_{|\beta} \xi^{\eta} + \begin{Bmatrix} \alpha \\ \tau\ \gamma \end{Bmatrix} \begin{Bmatrix} \tau \\ \beta\ \eta \end{Bmatrix} \xi^{\eta} \\ &\quad - \begin{Bmatrix} \alpha \\ \tau\ \beta \end{Bmatrix} \begin{Bmatrix} \tau \\ \gamma\ \eta \end{Bmatrix} \xi^{\eta} \\ &= \left[\begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix}_{|\gamma} - \begin{Bmatrix} \alpha \\ \gamma\ \eta \end{Bmatrix}_{|\beta} + \begin{Bmatrix} \alpha \\ \tau\ \gamma \end{Bmatrix} \begin{Bmatrix} \tau \\ \beta\ \eta \end{Bmatrix} \right. \\ &\quad \left. - \begin{Bmatrix} \alpha \\ \tau\ \beta \end{Bmatrix} \begin{Bmatrix} \tau \\ \gamma\ \eta \end{Bmatrix} \right] \xi^{\eta} \end{aligned} \tag{5.9}$$

The object in braces must be a tensor by the quotient theorem. It is known as the *Riemann curvature tensor*, and as we shall see, it plays a central role in the geometric structure of a Riemann space. We denote it by $R^{\alpha}{}_{\eta\beta\gamma}$:

$$R^{\alpha}{}_{\eta\beta\gamma} = \begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix}_{|\gamma} - \begin{Bmatrix} \alpha \\ \eta\ \gamma \end{Bmatrix}_{|\beta} + \begin{Bmatrix} \alpha \\ \tau\ \gamma \end{Bmatrix} \begin{Bmatrix} \tau \\ \beta\ \eta \end{Bmatrix} - \begin{Bmatrix} \alpha \\ \tau\ \beta \end{Bmatrix} \begin{Bmatrix} \tau \\ \gamma\ \eta \end{Bmatrix} \tag{5.10}$$

Note that, although we have introduced $R^{\alpha}{}_{\eta\beta\gamma}$ for a Riemann space, it is evident that our derivation holds also in a general affine space, since it involves only the coefficients of connection, and not the metric tensor itself. In the more general case, one need only replace $-\begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix}$ by $\Gamma^{\alpha}{}_{\beta\eta}$. However, as soon as we lower an index to form the tensor $R_{\alpha\eta\beta\gamma}$, we commit ourselves to a metric space.

In terms of the Riemann tensor, Eq. (5.9) can be written

$$\xi^{\alpha}{}_{\|\beta\|\gamma} - \xi^{\alpha}{}_{\|\gamma\|\beta} = R^{\alpha}{}_{\eta\beta\gamma}\xi^{\eta} \tag{5.11}$$

Thus the necessary condition (5.5) that a Riemann space admit a Lorentz metric can be written

$$R^{\alpha}{}_{\eta\beta\gamma}\xi^{\eta} = \xi^{\alpha}{}_{\|\beta\|\gamma} - \xi^{\alpha}{}_{\|\gamma\|\beta} = 0 \tag{5.12}$$

Since ξ^{η} is an arbitrary vector, this implies the following simple property, which is a necessary condition for a space to have a Lorentz metric:

$$R^{\alpha}{}_{\eta\beta\gamma} = 0 \tag{5.13}$$

For simplicity we term a space *flat* if its *Riemann curvature tensor vanishes*. Equation (5.13) then states that *a space with a Lorentz metric is flat*. The converse statement, that a (physically acceptable) space has a Lorentz metric if it is flat, will be proved in Sec. 5.6, after we have further investigated the properties of the Riemann tensor.

Equation (5.13) now represents *a field equation for flat and gravity-free space* since the Lorentz metric is certainly a solution. Furthermore, since it involves the first derivatives of the Christoffel symbols linearly, it is second-order and quasi-linear in the metric tensor. It thus satisfies the field-equation criteria (*b*) and (*d*). This special field equation for *gravity-free* space will be a very helpful guide in obtaining the general field equations of gravity in a *nonflat* space.

In the process of achieving the primary aim of this section, the necessary condition for a space to have a Lorentz metric, we have also obtained Eq. (5.11), which is interesting and important in its own right. It is also possible to obtain a similar result for higher-rank tensors. Indeed, for a tensor of the form $\xi^\alpha\eta^\delta$, we have, by the product rule for covariant derivatives,

$$(\xi^\alpha\eta^\delta)_{\|\beta} = \xi^\alpha{}_{\|\beta}\eta^\delta + \xi^\alpha\eta^\delta{}_{\|\beta} \tag{5.14}$$

and

$$(\xi^\alpha\eta^\delta)_{\|\beta\|\gamma} = \xi^\alpha{}_{\|\beta\|\gamma}\eta^\delta + \xi^\alpha\eta^\delta{}_{\|\beta\|\gamma} + \xi^\alpha{}_{\|\beta}\eta^\delta{}_{\|\gamma} + \xi^\alpha{}_{\|\gamma}\eta^\delta{}_{\|\beta} \tag{5.15}$$

Interchanging the indices β and γ gives

$$(\xi^\alpha\eta^\delta)_{\|\gamma\|\beta} = \xi^\alpha{}_{\|\gamma\|\beta}\eta^\delta + \xi^\alpha\eta^\delta{}_{\|\gamma\|\beta} + \xi^\alpha{}_{\|\gamma}\eta^\delta{}_{\|\beta} + \xi^\alpha{}_{\|\beta}\eta^\delta{}_{\|\gamma} \tag{5.16}$$

The difference of the two expressions is then

$$(\xi^\alpha\eta^\delta)_{\|\beta\|\gamma} - (\xi^\alpha\eta^\delta)_{\|\gamma\|\beta} = (\xi^\alpha{}_{\|\beta\|\gamma} - \xi^\alpha{}_{\|\gamma\|\beta})\eta^\delta + (\eta^\delta{}_{\|\beta\|\gamma} - \eta^\delta{}_{\|\gamma\|\beta})\xi^\alpha \tag{5.17}$$

Inserting (5.11) into (5.17), we have

$$(\xi^\alpha\eta^\delta)_{\|\beta\|\gamma} - (\xi^\alpha\eta^\delta)_{\|\gamma\|\beta} = R^\alpha{}_{\tau\beta\gamma}\xi^\tau\eta^\delta + R^\delta{}_{\tau\beta\gamma}\xi^\alpha\eta^\tau \tag{5.18}$$

which is the result analogous to (5.11). However, we know from Chap. 2 that *any* second-rank tensor can be written as a linear combination of such products. Thus, for any second-rank tensor $T^{\alpha\delta}$, we must have

$$T^{\alpha\delta}{}_{\|\beta\|\gamma} - T^{\alpha\delta}{}_{\|\gamma\|\beta} = R^\alpha{}_{\tau\beta\gamma}T^{\tau\delta} + R^\delta{}_{\tau\beta\gamma}T^{\alpha\tau} \tag{5.19}$$

The generalization of this process to yet higher rank tensors is evident.

Before ending this section we also wish to note the immediate consequence of (5.11):

$$\xi_{\alpha\|\beta\|\gamma} - \xi_{\alpha\|\gamma\|\beta} = R_{\alpha\rho\beta\gamma}\xi^\rho \tag{5.11$'$}$$

which is useful in covariant differentiation of covariant vectors. This formula makes sense only in a metric space, as we remarked after Eq. (5.10). On making the same calculations for the interchange of derivatives of covariant vectors, as we did for contravariant vectors, we would have obtained

$$\xi_{\alpha\|\beta\|\gamma} - \xi_{\alpha\|\gamma\|\beta} = -R^\eta{}_{\alpha\beta\gamma}\xi_\eta \tag{5.11$''$}$$

a formula valid in the general affine case.

5.3 Symmetry Properties of the Riemann Tensor

Unfortunately, the Riemann tensor is a rather cumbersome tensor, with 4^4, or 256, components. However, the number of *independent* components is much less than this because of various symmetry relations. By inspection of (5.10), we see that it is antisymmetric in the third and fourth indices, β and γ. Thus the $\beta\gamma$ subblock has only 6, instead of 16, *independent* components. Combined with the 16 components of the $\alpha\eta$ subblock, this reduces the number of independent components to at most 96.

Consider now an arbitrary vector field $\boldsymbol{\xi}$ and the square of its length

$$\varphi = g_{\tau\lambda}\xi^\tau\xi^\lambda \tag{5.20}$$

Since φ is a scalar, its ordinary and covariant derivatives are the same, so we have

$$\varphi_{\|\beta} = \varphi_{|\beta} \tag{5.21}$$

Antisymmetrizing $\varphi_{\|\lambda\|\tau}$, we have, by the results of Chap. 3,

$$\{\varphi_{|\beta|\gamma}\} = \varphi_{\|\beta\|\gamma} - \varphi_{\|\gamma\|\beta} = \varphi_{|\beta|\gamma} - \varphi_{|\gamma|\beta} = 0 \tag{5.22}$$

Now since

$$\varphi_{\|\beta} = g_{\tau\lambda}\xi^\tau{}_{\|\beta}\xi^\lambda + g_{\tau\lambda}\xi^\tau\xi^\lambda{}_{\|\beta} = 2g_{\tau\lambda}\xi^\tau\xi^\lambda{}_{\|\beta} \tag{5.23}$$

and

$$\varphi_{\|\beta\|\gamma} = 2g_{\tau\lambda}\xi^\tau{}_{\|\gamma}\xi^\lambda{}_{\|\beta} + 2g_{\tau\lambda}\xi^\tau\xi^\lambda{}_{\|\beta\|\gamma} \tag{5.24}$$

we have also

$$\varphi_{\|\beta\|\gamma} - \varphi_{\|\gamma\|\beta} = 2g_{\tau\lambda}\xi^{\tau}(\xi^{\lambda}{}_{\|\beta\|\gamma} - \xi^{\lambda}{}_{\|\gamma\|\beta}) = 0 \tag{5.25}$$

Using (5.11), this can be stated in terms of the Riemann tensor as

$$\varphi_{\|\beta\|\gamma} - \varphi_{\|\gamma\|\beta} = 2\xi_{\alpha}R^{\alpha}{}_{\eta\beta\gamma}\xi^{\eta} = 0 \tag{5.26}$$

Thus, by (5.22) and (5.26), we can assert, for an arbitrary vector field ξ, that

$$R_{\alpha\eta\beta\gamma}\xi^{\eta}\xi^{\alpha} = 0 \tag{5.27}$$

Let us now choose at a given point of our Riemann space the arbitrary vector ξ to be a unit vector with the α_0 component equal to 1 and all other components zero. Then, noting that α_0 is *not* a summation index, we have

$$R_{\alpha_0\alpha_0\beta\gamma} = 0 \tag{5.28}$$

That is, the diagonal terms of the subblock are zero. We can express the summation in (5.27) as

$$R_{\alpha\eta\beta\gamma}\xi^{\alpha}\xi^{\eta} = \tfrac{1}{2}(R_{\alpha\eta\beta\gamma} + R_{\eta\alpha\beta\gamma})\xi^{\alpha}\xi^{\eta} = 0 \tag{5.29}$$

Now we choose ξ to have two nonzero components, ξ^{μ_0} and ξ^{ν_0}, both equal to 1. Since the diagonal terms are zero according to (5.28), we have

$$R_{\mu_0\nu_0\beta\gamma} + R_{\nu_0\mu_0\beta\gamma} = 0 \tag{5.30}$$

Since μ_0 and ν_0 are any two components, we have the resultant antisymmetry relation

$$R_{\alpha\eta\beta\gamma} = -R_{\eta\alpha\beta\gamma} \tag{5.31}$$

That is, the Riemann tensor is antisymmetric in the first and second indices. We now have at most 6 independent components in the $\alpha\eta$ subblock, so together with the 6 in the $\beta\gamma$ subblock, there remain at most 36 independent components.

To obtain the final symmetry property, consider an antisymmetrized tensor formed from an arbitrary vector ξ_α:

$$\{\xi_{\alpha\|\beta\|\gamma} - \xi_{\beta\|\alpha\|\gamma}\} = \{\{\xi_{\alpha\|\beta}\}_{\|\gamma}\} \tag{5.32}$$

By the general property of antisymmetrized tensors (Chap. 3), we may replace the covariant derivatives by ordinary derivatives, so we have

$$\{\xi_{\alpha\|\beta\|\gamma} - \xi_{\beta\|\alpha\|\gamma}\} = \{\{\xi_{\alpha|\beta}\}_{|\gamma}\} \tag{5.33}$$

Furthermore, we know that an exact antisymmetric tensor is closed, so that the repeated application of differentiation with antisymmetrization always produces a null tensor, and the above expression must be identically zero:

$$\{\xi_{\alpha\|\beta\|\gamma} - \xi_{\beta\|\alpha\|\gamma}\} = 0 \tag{5.34}$$

In the foregoing we have repeated the same considerations which led to the preceding antisymmetry law (5.31); only we started with the vector ξ_α instead of the scalar φ. On the other hand, the indices of the second term in the above null expression may be cyclically permuted without changing the expression. This gives

$$\{\xi_{\alpha\|\beta\|\gamma} - \xi_{\alpha\|\gamma\|\beta}\} = 0 \tag{5.35}$$

By virtue of Eq. (5.11) we may write this as

$$\{\xi_{\alpha\|\beta\|\gamma} - \xi_{\alpha\|\gamma\|\beta}\} = \{R_{\alpha\eta\beta\gamma}\xi^{\eta}\} = 0 \tag{5.36}$$

It is convenient to introduce here the notational convention

$$\{R_{\alpha\eta\beta\gamma}\xi^{\eta}\}_{(\alpha,\beta,\gamma)} = 0 \tag{5.37}$$

which means that only the indices α, β, and γ are included in the antisymmetrization, not the summation index η. Since η does not enter the antisymmetrization process, we can take ξ^η outside the brackets.

$$\{R_{\alpha\eta\beta\gamma}\xi^{\eta}\}_{(\alpha,\beta,\gamma)} = \{R_{\alpha\eta\beta\gamma}\}_{(\alpha,\beta,\gamma)}\xi^{\eta} = 0 \tag{5.38}$$

Thus, since ξ^η is an arbitrary vector, we have

$$\{R_{\alpha\eta\beta\gamma}\}_{(\alpha,\beta,\gamma)} = 0 \tag{5.39}$$

Relabeling the indices, we have also the relation

$$\{R_{\eta\alpha\beta\gamma}\}_{(\eta,\beta,\gamma)} = 0 \tag{5.40}$$

Let us write out these last two relations, (5.39) and (5.40); making use of the antisymmetry in the first and second indices and in the third and fourth indices, we obtain

$$R_{\alpha\eta\beta\gamma} + R_{\beta\eta\gamma\alpha} + R_{\gamma\eta\alpha\beta} = 0 \tag{5.41}$$

$$R_{\eta\alpha\beta\gamma} + R_{\beta\alpha\gamma\eta} + R_{\gamma\alpha\eta\beta} = 0 \tag{5.42}$$

Subtracting (5.42) from (5.41), remembering the antisymmetry in the first two indices, we get

$$2R_{\alpha\eta\beta\gamma} + R_{\beta\eta\gamma\alpha} + R_{\gamma\eta\alpha\beta} - R_{\beta\alpha\gamma\eta} - R_{\gamma\alpha\eta\beta} = 0 \tag{5.43}$$

or equivalently,

$$2R_{\alpha\eta\beta\gamma} - R_{\eta\beta\gamma\alpha} - R_{\eta\gamma\alpha\beta} - R_{\alpha\beta\eta\gamma} - R_{\alpha\gamma\beta\eta} = 0 \tag{5.44}$$

Regrouping terms, we can write this as

$$2R_{\alpha\eta\beta\gamma} - (R_{\eta\beta\gamma\alpha} + R_{\alpha\beta\eta\gamma}) - (R_{\eta\gamma\alpha\beta} + R_{\alpha\gamma\beta\eta}) = 0 \tag{5.45}$$

This can be simplified if we relabel the indices in (5.41) and (5.42) to give the pair of relations

$$R_{\eta\beta\gamma\alpha} + R_{\alpha\beta\eta\gamma} = -R_{\gamma\beta\alpha\eta} \tag{5.46}$$

$$R_{\eta\gamma\alpha\beta} + R_{\alpha\gamma\beta\eta} = -R_{\beta\gamma\eta\alpha} \tag{5.47}$$

which we then may substitute in (5.45) to get

$$2R_{\alpha\eta\beta\gamma} + R_{\gamma\beta\alpha\eta} + R_{\beta\gamma\eta\alpha} = 0 \tag{5.48}$$

But the Riemann tensor is antisymmetric in the first two and in the last two indices, so the last two terms of (5.48) are equal. Thus

$$2(R_{\alpha\eta\beta\gamma} + R_{\gamma\beta\alpha\eta}) = 2(R_{\alpha\eta\beta\gamma} - R_{\beta\gamma\alpha\eta}) = 0 \tag{5.49}$$

We then obtain, finally,

$$R_{\alpha\eta\beta\gamma} = R_{\beta\gamma\alpha\eta} \tag{5.50}$$

which is our final symmetry property. Thus we see that $R_{\alpha\eta\beta\gamma}$ is *antisymmetric* in each of the index pairs $\alpha\eta$ and $\beta\gamma$, but is *symmetric* under interchange of the pairs. If we treat the first index pair as one index capable of assuming six values, and similarly for the second pair, we take

account of the antisymmetry in those pairs. Then the symmetry under interchange of the pairs implies that we have left at most the same number of independent components as a symmetric 6×6 matrix, that is, 21 independent components. Let us repeat in summary the symmetry properties we have obtained:

$$\begin{aligned} R_{\alpha\eta\beta\gamma} &= -R_{\alpha\eta\gamma\beta} \\ R_{\alpha\eta\beta\gamma} &= -R_{\eta\alpha\beta\gamma} \\ R_{\alpha\eta\beta\gamma} &= R_{\beta\gamma\alpha\eta} \end{aligned} \tag{5.51}$$

There is, actually, one more symmetry property contained in (5.39) which is not contained in the set (5.51) and which reduces the number of independent components ultimately to 20. The additional condition can be expressed as follows: Let us consider the component R_{0123} of the Riemann tensor. Under all possible permutations of the indices 0, 1, 2, 3, we obtain 4!, or 24, formally different components. However, when applying (5.51) to R_{0123}, the indices 0 and 1, on the one hand, and 2 and 3, on the other hand, remain adjacent, and we can identify all components as multiples of the basic three

$$R_{0123},\ R_{0231},\ R_{0312} \tag{5.52}$$

But from the cyclic symmetry (5.42) it follows that

$$R_{1023} + R_{2031} + R_{3012} = 0 \tag{5.53}$$

which establishes a relation among the components (5.52) and is therefore not contained in (5.51). On the other hand, (5.39) is important in its own right and will be useful in the following sections, so we repeat it also in our summary of symmetry properties:

$$\{R_{\alpha\eta\beta\gamma}\}_{(\alpha,\beta,\gamma)} = 0 \tag{5.54}$$

5.4 The Bianchi Identities

In the preceding section we obtained several useful algebraic symmetry relations for the Riemann curvature tensor $R_{\alpha\eta\beta\gamma}$, which are summarized in (5.51) and (5.54). In this section we shall use these relations to obtain a set of symmetry relations on the covariant derivatives of the Riemann tensor.

Lowering the index α in Eq. (5.11) gives

(5.55) $$\xi_{\alpha\|\beta\|\gamma} - \xi_{\alpha\|\gamma\|\beta} = R_{\alpha\eta\beta\gamma}\xi^\eta$$

Using the product rule for covariant derivatives, we then obtain

(5.56) $$(\xi_{\alpha\|\beta\|\gamma} - \xi_{\alpha\|\gamma\|\beta})_{\|\delta} = R_{\alpha\eta\beta\gamma\|\delta}\xi^\eta + R_{\alpha\eta\beta\gamma}\xi^\eta{}_{\|\delta}$$

Let us antisymmetrize this with respect to the indices β, γ, and δ:

(5.57) $$\{\xi_{\alpha\|\beta\|\gamma\|\delta} - \xi_{\alpha\|\gamma\|\beta\|\delta}\}_{(\beta,\gamma,\delta)} = \{R_{\alpha\eta\beta\gamma\|\delta}\}_{(\beta,\gamma,\delta)}\xi^\eta + \{R_{\alpha\eta\beta\gamma}\xi^\eta{}_{\|\delta}\}_{(\beta,\gamma,\delta)}$$

By the nature of the process of antisymmetrization, we may make an even permutation of the indices β, γ, and δ on the left side of the above equation to obtain

(5.58) $$\{(\xi_{\alpha\|\delta})_{\|\beta\|\gamma} - (\xi_{\alpha\|\delta})_{\|\gamma\|\beta}\}_{(\beta,\gamma,\delta)} = \{R_{\alpha\eta\beta\gamma\|\delta}\}_{(\beta,\gamma,\delta)}\xi^\eta + \{R_{\alpha\eta\beta\gamma}\xi^\eta{}_{\|\delta}\}_{(\beta,\gamma,\delta)}$$

Note that the tensor $\xi_{\alpha\|\delta}$ appears in both terms of the left side. Applying Eq. (5.19) to this tensor, we have

(5.59) $$(\xi_{\alpha\|\delta})_{\|\beta\|\gamma} - (\xi_{\alpha\|\delta})_{\|\gamma\|\beta} = R_{\alpha\eta\beta\gamma}\xi^\eta{}_{\|\delta} + R_{\delta\eta\beta\gamma}\xi_\alpha{}^{\|\eta}$$

where $\xi_\alpha{}^{\|\eta} = g^{\eta\nu}\xi_{\alpha\|\nu}$ is the contravariant form of $\xi_{\alpha\|\eta}$. Antisymmetrizing, we obtain an alternative form for the left side of (5.58).

(5.60) $$\{(\xi_{\alpha\|\delta})_{\|\beta\|\gamma} - (\xi_{\alpha\|\delta})_{\|\gamma\|\beta}\}_{(\beta,\gamma,\delta)} = \{R_{\alpha\eta\beta\gamma}\xi^\eta{}_{\|\delta}\}_{(\beta,\gamma,\delta)} + \{R_{\delta\eta\beta\gamma}\}_{(\beta,\gamma,\delta)}\xi_\alpha{}^{\|\eta}$$

Comparing (5.58) and (5.60), we then have

(5.61) $$\{R_{\alpha\eta\beta\gamma\|\delta}\}_{(\beta,\gamma,\delta)}\xi^\eta = \{R_{\delta\eta\beta\gamma}\}_{(\beta,\gamma,\delta)}\xi_\alpha{}^{\|\eta}$$

However, the right side of the above is zero by virtue of the symmetry relation (5.39), so since ξ^η is an arbitrary vector, we have finally

(5.62) $$\{R_{\alpha\eta\beta\gamma\|\delta}\}_{(\beta,\gamma,\delta)} = 0$$

The symmetry relations contained in (5.62) are known as the Bianchi identities. They will be quite useful when we investigate the gravitational field equations in Sec. 5.8.

One could expect a priori that there should exist numerous relations between the 20 components of the Riemann tensor. This is because it is constructed out of the 10 components of the metric tensor, and a large number of consistency conditions are clearly necessary in order that a tensor with the symmetries obtained in Sec. 5.3 be of such form.

5.5 Integrability and the Riemann Tensor

By our definition of parallel displacement, the change in the α component of a vector ξ^α under parallel displacement is given by

(5.63) $$d\xi^\alpha = -\begin{Bmatrix}\alpha \\ \beta\ \gamma\end{Bmatrix}\xi^\beta\,dx^\gamma$$

If we now are given a curve between two points and the value of ξ^α at one of the points, the above equation enables us to compute ξ^α at the other point; however, if we displace the same vector along a *different* curve connecting the same two points, we have no reason to expect to arrive at the same value of ξ^α at the endpoint. Equivalently, if we displace a vector parallel to itself around a *closed path*, we have no reason to expect the ξ^α to return to their initial values. We shall be concerned in this section with the question of how the change in parallel displaced vectors depends upon the path taken. That is, we shall study the dependence of the functional

(5.64) $$I^\alpha(\Gamma) = \int_\Gamma d\xi^\alpha$$

on the path Γ.

Let us first consider the simplest and most important case. We assume that we start with an arbitrary vector at an arbitrarily given point and construct from it by parallel displacement a uniquely determined vector field in some finite neighborhood. This is clearly possible only if the result of the parallel displacement is independent of the path used to reach the final point considered. We thereby generate a vector field:

(5.65) $$\xi^\alpha = \xi^\alpha(x^\mu)$$

The law of parallel displacement (5.63) then states that the derivatives of ξ^α are given by

(5.66) $$\xi^\alpha{}_{|\gamma} = \frac{\partial\xi^\alpha}{\partial x^\gamma} = -\begin{Bmatrix}\alpha \\ \beta\ \gamma\end{Bmatrix}\xi^\beta$$

Note that this implies that the covariant derivatives of ξ^α are zero. Furthermore, the order of ordinary differentiation of a vector field is irrelevant; that is,

(5.67) $$\xi^\alpha{}_{|\gamma|\delta} = \xi^\alpha{}_{|\delta|\gamma}$$

We see that we must have, according to (5.66),

(5.68)
$$\left(\left\{ {\alpha \atop \beta\;\gamma} \right\} \xi^\beta\right)_{|\delta} = \left(\left\{ {\alpha \atop \beta\;\delta} \right\} \xi^\beta\right)_{|\gamma}$$

This can be put in more meaningful form by applying the product rule for ordinary derivatives,

(5.69)
$$\left\{ {\alpha \atop \beta\;\gamma} \right\}_{|\delta} \xi^\beta + \left\{ {\alpha \atop \beta\;\gamma} \right\} \xi^\beta{}_{|\delta} = \left\{ {\alpha \atop \beta\;\delta} \right\}_{|\gamma} \xi^\beta + \left\{ {\alpha \atop \beta\;\delta} \right\} \xi^\beta{}_{|\gamma}$$

and substituting for the first derivatives of ξ^β from (5.66). We obtain

(5.70)
$$\left\{ {\alpha \atop \beta\;\gamma} \right\}_{|\delta} \xi^\beta - \left\{ {\alpha \atop \beta\;\gamma} \right\} \left\{ {\beta \atop \tau\;\delta} \right\} \xi^\tau = \left\{ {\alpha \atop \beta\;\delta} \right\}_{|\gamma} \xi^\beta - \left\{ {\alpha \atop \beta\;\delta} \right\} \left\{ {\beta \atop \tau\;\gamma} \right\} \xi^\tau$$

Rearranging terms and relabeling indices, we then get

(5.71)
$$\left[\left\{ {\alpha \atop \beta\;\gamma} \right\}_{|\delta} - \left\{ {\alpha \atop \beta\;\delta} \right\}_{|\gamma} + \left\{ {\alpha \atop \tau\;\delta} \right\} \left\{ {\tau \atop \beta\;\gamma} \right\} - \left\{ {\alpha \atop \tau\;\gamma} \right\} \left\{ {\tau \atop \beta\;\delta} \right\}\right] \xi^\beta = 0$$

or by definition of the Riemann tensor,

(5.72)
$$R^\alpha{}_{\beta\gamma\delta}\xi^\beta = 0$$

Since ξ^β can be arbitrarily chosen at any single given point, we must then have

(5.73)
$$R^\alpha{}_{\beta\gamma\delta} = 0 \qquad \text{(condition for integrability)}$$

Thus we see that a space must be flat (have a null Riemann curvature tensor) if we are able to establish in it a vector field by parallel displacement, starting with an arbitrary value at an arbitrary point of the space.

Let us next investigate the case of a space whose Riemann curvature tensor is not necessarily zero. We shall study the effect of a nonzero Riemann tensor on the parallel displacement of a vector. Consider two paths between some initial point P_i and a final point P_f. One path consists of a displacement along the vector $\mathbf{dx}$, followed by a displacement along $\mathbf{d\hat{x}}$, and the other is along the same vectors in reverse order as illustrated in Fig. 5.1. We shall compute the change of ξ^α along

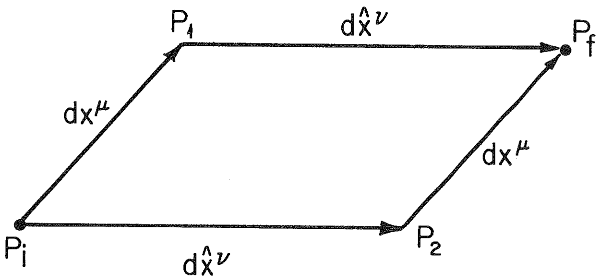


Fig. 5.1

both paths and compare the results. By the law of parallel displacement, the change of ξ^α between P_i and P_1 (along the first path) is

(5.74)
$$d\xi^\alpha(P_i,P_1) = -\left\{ {\alpha \atop \eta\;\rho} \right\} \xi^\rho\, dx^\eta$$

(Unless explicitly noted, the vector ξ^α and the Christoffel symbols are always evaluated at P_i as in the right side above.) At P_1 the displaced vector is then given by

(5.75)
$$\xi^\alpha(P_1) = \xi^\alpha - \left\{ {\alpha \atop \eta\;\rho} \right\} \xi^\rho\, dx^\eta$$

The Christoffel symbols at P_1 are, to first order in a Taylor series expansion in the vector $\mathbf{dx}$,

(5.76)
$$\left\{ {\alpha \atop \gamma\;\beta} \right\}_{P_1} = \left\{ {\alpha \atop \gamma\;\beta} \right\} + \left\{ {\alpha \atop \gamma\;\beta} \right\}_{|\eta} dx^\eta$$

Next applying the law of parallel displacement to the vector $\xi^\alpha(P_1)$, using the Christoffel symbols at P_1 given by (5.76), we obtain the change in ξ^α between P_1 and P_f to second order in the displacement vectors $\mathbf{dx}$ and $\mathbf{d\hat{x}}$:

(5.77)
$$d\xi^\alpha(P_1,P_f) = -\left[\left\{ {\alpha \atop \gamma\;\beta} \right\} + \left\{ {\alpha \atop \gamma\;\beta} \right\}_{|\eta} dx^\eta\right]\left[\xi^\beta - \left\{ {\beta \atop \eta\;\rho} \right\} \xi^\rho\, dx^\eta\right] d\hat{x}^\gamma$$

Rearranging terms, we can write this in the form

(5.78)
$$d\xi^\alpha(P_1,P_f) = -\left\{ {\alpha \atop \gamma\;\beta} \right\} \xi^\beta\, d\hat{x}^\gamma - \left\{ {\alpha \atop \gamma\;\beta} \right\}_{|\eta} \xi^\beta\, dx^\eta\, d\hat{x}^\gamma + \left\{ {\alpha \atop \gamma\;\beta} \right\} \left\{ {\beta \atop \eta\;\rho} \right\} \xi^\rho\, dx^\eta\, d\hat{x}^\gamma$$

Again, as we noted above, the terms on the right side are evaluated at P_i. Thus the value of ξ^α after traversing the entire path is, to second order in the displacement vectors $\mathbf{dx}$ and $\mathbf{d\hat{x}}$,

$$(5.79)\quad \xi^\alpha(P_iP_1P_f) = \xi^\alpha + d\xi^\alpha(P_iP_1) + d\xi^\alpha(P_1P_f)$$
$$= \xi^\alpha - \begin{Bmatrix} \alpha \\ \eta\;\beta \end{Bmatrix} \xi^\beta\, dx^\eta - \begin{Bmatrix} \alpha \\ \gamma\;\beta \end{Bmatrix} \xi^\beta\, d\hat{x}^\gamma - \begin{Bmatrix} \alpha \\ \gamma\;\beta \end{Bmatrix}_{|\eta} \xi^\beta\, dx^\eta\, d\hat{x}^\gamma + \begin{Bmatrix} \alpha \\ \gamma\;\lambda \end{Bmatrix} \begin{Bmatrix} \lambda \\ \eta\;\beta \end{Bmatrix} \xi^\beta\, dx^\eta\, d\hat{x}^\gamma$$

The result along the second path, $P_i \to P_2 \to P_f$, is gotten by simply interchanging $\mathbf{dx}$ and $\mathbf{d\hat{x}}$:

$$(5.80)\quad \xi^\alpha(P_iP_2P_f) = \xi^\alpha - \begin{Bmatrix} \alpha \\ \gamma\;\beta \end{Bmatrix} \xi^\beta\, d\hat{x}^\gamma - \begin{Bmatrix} \alpha \\ \eta\;\beta \end{Bmatrix} \xi^\beta\, dx^\eta - \begin{Bmatrix} \alpha \\ \eta\;\beta \end{Bmatrix}_{|\gamma} \xi^\beta\, d\hat{x}^\gamma\, dx^\eta + \begin{Bmatrix} \alpha \\ \eta\;\lambda \end{Bmatrix} \begin{Bmatrix} \lambda \\ \gamma\;\beta \end{Bmatrix} \xi^\beta\, d\hat{x}^\gamma\, dx^\eta$$

The difference between the ξ^α obtained by parallel displacement along the two routes is therefore

$$(5.81)\quad \Delta\xi^\alpha = \begin{Bmatrix} \alpha \\ \eta\;\beta \end{Bmatrix}_{|\gamma} \xi^\beta\, d\hat{x}^\gamma\, dx^\eta - \begin{Bmatrix} \alpha \\ \gamma\;\beta \end{Bmatrix}_{|\eta} \xi^\beta\, d\hat{x}^\gamma\, dx^\eta + \begin{Bmatrix} \alpha \\ \gamma\;\lambda \end{Bmatrix} \begin{Bmatrix} \lambda \\ \eta\;\beta \end{Bmatrix} \xi^\beta\, dx^\eta\, d\hat{x}^\gamma - \begin{Bmatrix} \alpha \\ \eta\;\lambda \end{Bmatrix} \begin{Bmatrix} \lambda \\ \gamma\;\beta \end{Bmatrix} \xi^\beta\, dx^\eta\, d\hat{x}^\gamma$$

By definition of the Riemann tensor, this is precisely

$$(5.82)\qquad \Delta\xi^\alpha = R^\alpha{}_{\beta\eta\gamma}\xi^\beta\, dx^\eta\, d\hat{x}^\gamma$$

Thus the value of ξ^α at the nearby point is independent of path if and only if $R^\alpha{}_{\beta\eta\gamma} = 0$, and for a nonzero Riemann tensor the difference in final value is given by (5.82).

We may avoid the use of too many differentials while studying the integrability conditions by use of the methods of calculus of variations. Indeed, let ξ^β be an arbitrary vector field and consider the integral

$$(5.83)\qquad J^\alpha[C] = -\int_C \begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix} \xi^\beta\, dx^\gamma$$

along a curve C which connects two given points in our Riemann space, say, P_i and P_f. In general, this integral will be a function of the curve C; we wish to find out for which vector fields this integral is path-independent. We perform a variation δx^γ which vanishes at the endpoints of C and find

$$(5.84)\quad \delta J^\alpha = -\int_C \left[\begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix}_{|\rho} \xi^\beta\, \delta x^\rho\, dx^\gamma + \begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix} \xi^\beta{}_{|\rho}\, \delta x^\rho\, dx^\gamma + \begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix} \xi^\beta \delta\, dx^\gamma \right]$$

We then integrate by parts over the last term and use the fact that $\delta x^\gamma = 0$ at the endpoints; we obtain, by use of definition (5.10),

$$(5.85)\quad \delta J^\alpha = \int_C \left[\begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix}_{|\rho} \xi^\beta + \begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix} \xi^\beta{}_{|\rho} \right] (\delta x^\gamma\, dx^\rho - \delta x^\rho\, dx^\gamma)$$
$$= \int_C \left[R^\alpha{}_{\beta\gamma\rho}\xi^\beta + \begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix} \xi^\beta{}_{\|\rho} - \begin{Bmatrix} \alpha \\ \beta\;\rho \end{Bmatrix} \xi^\beta{}_{\|\gamma} \right] \delta x^\gamma\, dx^\rho$$

We may even choose ξ^β in such a way that on C all its covariant derivatives vanish. In this case,

$$(5.86)\qquad \delta J^\alpha = \int_C R^\alpha{}_{\beta\gamma\rho}\xi^\beta\, \delta x^\gamma\, dx^\rho$$

which is indeed just the integral form of (5.82).

Let us assume, on the other hand, that the Riemann tensor vanishes identically. We may define a vector field by the system of linear partial differential equations

$$(5.87)\qquad \xi^\alpha{}_{|\gamma} = -\begin{Bmatrix} \alpha \\ \beta\;\gamma \end{Bmatrix} \xi^\beta$$

The condition for integrability of this system was shown above to be the vanishing of the Riemann tensor, which is now fulfilled. It follows from the general theory of such differential systems that this condition is also sufficient for the existence of a solution. Hence $R_{\alpha\beta\gamma\delta} = 0$ indeed implies the existence of a vector field with zero covariant derivative $\xi^\alpha{}_{\|\gamma} = \mathbf{0}$. We may prescribe the initial values of the vector ξ^α at one point arbitrarily and continue this vector into the neighborhood by means of (5.87).

Let us now summarize the results of this section. We can establish a vector field $\xi^\alpha(x)$ by parallel displacement of an arbitrary vector ξ^α from some initial point to all points in a neighborhood in the Riemann space (by an arbitrary route) if and only if the Riemann tensor of the space is identically zero. In other words, we have found that (5.87) is integrable if and only if the space is flat, that is, has a zero Riemann curvature tensor. We shall henceforth refer to such a space as an integrable space. In addition, note that the covariant derivative of a vector field formed by parallel displacement is clearly zero. Therefore such a vector field is a natural generalization of a constant vector field in Cartesian coordinates. The results of this section then indicate that such a generalized constant vector field can exist only in a flat space with a zero Riemann curvature tensor.

5.6 Pseudo-Euclidean and Flat Spaces

From the previous section we know that we can establish a vector field with a zero covariant derivative if and only if the Riemann tensor is everywhere zero. We shall show in this section that the existence of such a generalized constant vector field ensures the existence of a coordinate system where the metric tensor has constant components. If one can find such a coordinate system where the metric tensor has constant components, the space is termed by definition a pseudo-Euclidean space. Thus we can say that the goal of this section is to show that a flat space (a space with a null Riemann tensor) is also pseudo-Euclidean.

Let us then suppose that the Riemann tensor of a space is everywhere zero; in that case we can establish a generalized constant vector field by the parallel displacement of some arbitrary vector ξ_α from an initial point to any given point in space. Let us do this with the following set of four vectors $\xi_\alpha^{(\gamma)}$:

(5.88)
$$\xi_\alpha^{(0)} = (1,0,0,0) \qquad \xi_\alpha^{(1)} = (0,1,0,0)$$
$$\xi_\alpha^{(2)} = (0,0,1,0) \qquad \xi_\alpha^{(3)} = (0,0,0,1)$$

which we can write more simply as

(5.89)
$$\xi_\alpha^{(\gamma)} = \delta_\alpha^{(\gamma)} = \begin{cases} 1 & \text{for } \gamma = \alpha \\ 0 & \text{for } \gamma \neq \alpha \end{cases}$$

(Note that γ is *not* a tensor index.) These vectors then represent the value of four generalized constant vector fields at an initial point, say, P_0, in some fixed but arbitrary coordinate system x^α:

(5.90)
$$\xi_\alpha^{(\gamma)}(P_0) = \delta_\alpha^{(\gamma)}$$

Now since each vector field $\xi_\alpha^{(\gamma)}(x)$ has a zero covariant derivative,

(5.91)
$$\xi_{\alpha\|\beta}^{(\gamma)} = 0$$

(we omit the argument x for clarity), it must also have a zero curl:

(5.92)
$$\xi_{\alpha\|\beta}^{(\gamma)} - \xi_{\beta\|\alpha}^{(\gamma)} = \xi_{\alpha|\beta}^{(\gamma)} - \xi_{\beta|\alpha}^{(\gamma)} = 0$$

The zero curl implies that each $\xi_\alpha^{(\gamma)}$ then has a scalar potential $\varphi^{(\gamma)}(x^\mu)$ such that

(5.93)
$$\xi_\alpha^{(\gamma)} = \frac{\partial \varphi^{(\gamma)}}{\partial x^\alpha}$$

Let us now use these scalar potential functions to define a transformation to a new coordinate system $\bar{x}^\gamma$:

(5.94)
$$\bar{x}^\gamma = \varphi^{(\gamma)}(x^\mu)$$

Such a transformation is permissible in *some neighborhood* of P_0 since the Jacobian of the transformation at P_0 is

(5.95)
$$\left\| \frac{\partial \varphi^{(\gamma)}}{\partial x^\alpha} \right\| = \|\xi_\alpha^{(\gamma)}\| = 1$$

In the following development we shall consider only the neighborhood of P_0 where the Jacobian (5.95) remains positive until we note otherwise. In the barred system Eq. (5.93) takes the form

(5.96)
$$\bar{\xi}_\alpha^{(\gamma)} = \frac{\partial \varphi^{(\gamma)}}{\partial \bar{x}^\alpha}$$

which, by virtue of the transformation (5.94), gives

(5.97)
$$\bar{\xi}_\alpha^{(0)} = (1,0,0,0) \qquad \bar{\xi}_\alpha^{(1)} = (0,1,0,0)$$
$$\bar{\xi}_\alpha^{(2)} = (0,0,1,0) \qquad \bar{\xi}_\alpha^{(3)} = (0,0,0,1)$$

point arbitrarily and continue this vector into the neighborhood by means of (5.87).

Let us now summarize the results of this section. We can establish a vector field $\xi^\alpha(x)$ by parallel displacement of an arbitrary vector ξ^α from some initial point to all points in a neighborhood in the Riemann space (by an arbitrary route) if and only if the Riemann tensor of the space is identically zero. In other words, we have found that (5.87) is integrable if and only if the space is flat, that is, has a zero Riemann curvature tensor. We shall henceforth refer to such a space as an integrable space. In addition, note that the covariant derivative of a vector field formed by parallel displacement is clearly zero. Therefore such a vector field is a natural generalization of a constant vector field in Cartesian coordinates. The results of this section then indicate that such a generalized constant vector field can exist only in a flat space with a zero Riemann curvature tensor.

5.6 Pseudo-Euclidean and Flat Spaces

From the previous section we know that we can establish a vector field with a zero covariant derivative if and only if the Riemann tensor is everywhere zero. We shall show in this section that the existence of such a generalized constant vector field ensures the existence of a coordinate system where the metric tensor has constant components. If one can find such a coordinate system where the metric tensor has constant components, the space is termed by definition a pseudo-Euclidean space. Thus we can say that the goal of this section is to show that a flat space (a space with a null Riemann tensor) is also pseudo-Euclidean.

Let us then suppose that the Riemann tensor of a space is everywhere zero; in that case we can establish a generalized constant vector field by the parallel displacement of some arbitrary vector ξ_α from an initial point to any given point in space. Let us do this with the following set of four vectors $\xi_\alpha^{(\gamma)}$:

$$\begin{aligned} &\xi_\alpha^{(0)} = (1,0,0,0) \qquad \xi_\alpha^{(1)} = (0,1,0,0) \\ &\xi_\alpha^{(2)} = (0,0,1,0) \qquad \xi_\alpha^{(3)} = (0,0,0,1) \end{aligned} \tag{5.88}$$

which we can write more simply as

$$\xi_\alpha^{(\gamma)} = \delta_\alpha^{(\gamma)} = \begin{cases} 1 & \text{for } \gamma = \alpha \\ 0 & \text{for } \gamma \neq \alpha \end{cases} \tag{5.89}$$

(Note that γ is *not* a tensor index.) These vectors then represent the value of four generalized constant vector fields at an initial point, say, P_0, in some fixed but arbitrary coordinate system x^α:

$$\xi_\alpha^{(\gamma)}(P_0) = \delta_\alpha^{(\gamma)} \tag{5.90}$$

Now since each vector field $\xi_\alpha^{(\gamma)}(x)$ has a zero covariant derivative,

$$\xi^{(\gamma)}_{\alpha\|\beta} = 0 \tag{5.91}$$

(we omit the argument x for clarity), it must also have a zero curl:

$$\xi^{(\gamma)}_{\alpha\|\beta} - \xi^{(\gamma)}_{\beta\|\alpha} = \xi^{(\gamma)}_{\alpha|\beta} - \xi^{(\gamma)}_{\beta|\alpha} = 0 \tag{5.92}$$

The zero curl implies that each $\xi_\alpha^{(\gamma)}$ then has a scalar potential $\varphi^{(\gamma)}(x^\mu)$ such that

$$\xi_\alpha^{(\gamma)} = \frac{\partial \varphi^{(\gamma)}}{\partial x^\alpha} \tag{5.93}$$

Let us now use these scalar potential functions to define a transformation to a new coordinate system $\bar{x}^\gamma$:

$$\bar{x}^\gamma = \varphi^{(\gamma)}(x^\mu) \tag{5.94}$$

Such a transformation is permissible in *some neighborhood* of P_0 since the Jacobian of the transformation at P_0 is

$$\left\| \frac{\partial \varphi^{(\gamma)}}{\partial x^\alpha} \right\| = \|\xi_\alpha^{(\gamma)}\| = 1 \tag{5.95}$$

In the following development we shall consider only the neighborhood of P_0 where the Jacobian (5.95) remains positive until we note otherwise. In the barred system Eq. (5.93) takes the form

$$\bar{\xi}_\alpha^{(\gamma)} = \frac{\partial \varphi^{(\gamma)}}{\partial \bar{x}^\alpha} \tag{5.96}$$

which, by virtue of the transformation (5.94), gives

$$\begin{aligned} &\bar{\xi}_\alpha^{(0)} = (1,0,0,0) \qquad \bar{\xi}_\alpha^{(1)} = (0,1,0,0) \\ &\bar{\xi}_\alpha^{(2)} = (0,0,1,0) \qquad \bar{\xi}_\alpha^{(3)} = (0,0,0,1) \end{aligned} \tag{5.97}$$

so each of the vectors $\bar{\xi}^{(\gamma)}_{\alpha}$ has *constant components everywhere* in the barred coordinate system $\bar{x}^{\alpha}$.

Consider next the set of 16 scalar inner products

$$\bar{\xi}^{(\gamma)}_{\alpha}\bar{\xi}^{(\delta)}_{\beta}\bar{g}^{\alpha\beta} = \xi^{(\gamma)}_{\alpha}\xi^{(\delta)}_{\beta}g^{\alpha\beta} \tag{5.98}$$

Since the vectors $\bar{\xi}^{(\gamma)}_{\alpha}$ and the metric tensor $\bar{g}^{\alpha\beta}$ have zero covariant derivatives, each of these scalars has a zero derivative:

$$\begin{aligned}(\bar{\xi}^{(\gamma)}_{\alpha}\bar{\xi}^{(\delta)}_{\beta}\bar{g}^{\alpha\beta})_{|\lambda} &= (\bar{\xi}^{(\gamma)}_{\alpha}\bar{\xi}^{(\delta)}_{\beta}\bar{g}^{\alpha\beta})_{\|\lambda} \\ &= \bar{\xi}^{(\gamma)}_{\alpha}\bar{\xi}^{(\delta)}_{\beta}\bar{g}^{\alpha\beta}{}_{\|\lambda} + \bar{\xi}^{(\gamma)}_{\alpha}\bar{\xi}^{(\delta)}_{\beta\|\lambda}\bar{g}^{\alpha\beta} + \bar{\xi}^{(\gamma)}_{\alpha\|\lambda}\bar{\xi}^{(\delta)}_{\beta}\bar{g}^{\alpha\beta} \\ &= 0 \qquad \text{for all pairs } \gamma \text{ and } \delta\end{aligned} \tag{5.99}$$

and is therefore a constant:

$$(\bar{\xi}^{(\gamma)}_{\alpha}\bar{\xi}^{(\delta)}_{\beta}\bar{g}^{\alpha\beta}) = \text{const} \tag{5.100}$$

From the fact that this inner product is a constant and from the explicit form of the constant vectors $\bar{\xi}^{(\gamma)}_{\alpha}$ given in (5.97), we see that each component of the metric tensor must be a constant:

$$\bar{\xi}^{(\gamma)}_{\alpha}\bar{\xi}^{(\delta)}_{\beta}\bar{g}^{\alpha\beta} = \bar{g}^{\gamma\delta} = \text{const} \tag{5.101}$$

We have therefore shown that, in the neighborhood of P_0, where the Jacobian (5.95) is nonzero, there is a coordinate system in which the metric tensor has constant components; the space is by definition pseudo-Euclidean.

The problem now remains to extend our analysis to include all space instead of the neighborhood of P_0, which we considered in the preceding paragraphs. As an aid in this extension, let us recall a few facts concerning the theory of symmetric matrices:

1. For any symmetric matrix G there is a nonsingular matrix A (whose transpose is denoted by A^T) which will transform G by a *congruence transformation* AGA^T to a diagonal matrix of the general form that is, with l diagonal elements equal to 1, m elements equal to -1, and n elements equal to zero. This is the well-known *Sylvester canonical form* for congruence transformations.

2. Although the matrix A in the transformation (5.102) is not unique, the set of diagonal elements in the Sylvester canonical form for any given

$$\begin{pmatrix} 1 &&&&&&&&&& \\ & \cdot &&&&&&&&& \\ && \cdot &&&&&&&& \\ &&& \cdot &&&&&&& \\ &&&& 1 &&&&&& \\ &&&&& -1 &&&&& \\ &&&&&& \cdot &&&& \\ &&&&&&& \cdot &&& \\ &&&&&&&& \cdot && \\ &&&&&&&&& -1 & \\ &&&&&&&&&& 0 \\ &&&&&&&&&&& \cdot \\ &&&&&&&&&&&& \cdot \\ &&&&&&&&&&&&& \cdot \\ &&&&&&&&&&&&&& 0 \end{pmatrix} = \begin{pmatrix} I_l && \\ & -I_m & \\ && 0_n \end{pmatrix} = AGA^T \tag{5.102}$$

matrix G is unique. This set of elements, composed of $+1$, -1, and 0, is termed the *signature* of the matrix G.

3. Any two matrices, say, G and H, which have the same signature and thereby the same canonical form, are related by a congruence transformation; that is, there exists a nonsingular matrix B such that $G = BHB^T$. (This last fact is indeed evident from facts 1 and 2.)

The use of these ideas in extending our preceding analysis over all space is quite straightforward, and we shall only briefly sketch the procedure. We have shown that in *some* neighborhood of *any* point P_0 there exists a transformation to a barred coordinate system in which the metric tensor is constant:

$$\bar{g}_{\alpha\beta} = \frac{\partial x^{\mu}}{\partial \bar{x}^{\alpha}}\frac{\partial x^{\nu}}{\partial \bar{x}^{\beta}}g_{\mu\nu} = \text{const} \tag{5.103}$$

By defining the matrices $(A)_{\alpha\mu} = \partial x^{\mu}/\partial\bar{x}^{\alpha}$, $(G)_{\mu\nu} = g_{\mu\nu}$, and $(\bar{G})_{\alpha\beta} = \bar{g}_{\alpha\beta}$, we can write this very simply in matrix notation as

$$\bar{G} = AGA^T = \text{const} \tag{5.104}$$

As we noted above, both $\bar{G}$ and G are assumed to have signature

$$(1,-1,-1,-1)$$

Indeed, it is evident that, without loss of generality, we may suppose that $\bar{G}$ is itself the Sylvester canonical form identical to the Lorentz metric.

Consider now a nearby point P_1 such that the neighborhoods of P_1 and P_0 overlap where the transformation (5.93) has a nonzero Jacobian

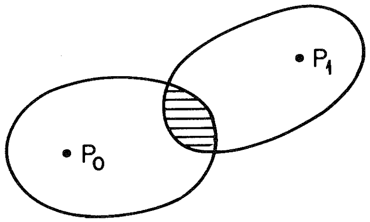


Fig. 5.2

(Fig. 5.2). The relevant φ functions for the neighborhoods need not necessarily agree, but we shall have a matrix equation similar to (5.104) in the neighborhood of P_1

$$\bar{G} = BGB^T = \text{const} \tag{5.105}$$

where $\bar{G}$ is again the Lorentz canonical form. In the *overlap* region we then have

$$\bar{G} = BGB^T = AGA^T = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \tag{5.106}$$

Since A and B are nonsingular matrices, this gives

$$G = B^{-1}\bar{G}(B^T)^{-1} = A^{-1}\bar{G}(A^T)^{-1} \tag{5.107}$$

Thus

$$\bar{G} = AB^{-1}\bar{G}(B^T)^{-1}A^T = (AB^{-1})\bar{G}(AB^{-1})^T \tag{5.108}$$

that is, AB^{-1} *transforms the Lorentz metric into itself* at every point in the overlap region. All matrices L such that

$$L\begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}L^T = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \tag{5.109}$$

form a group, the so-called *Lorentz group* of matrices or of linear transformations, which is well known in the special theory of relativity. By

definition, AB^{-1} then belongs to the Lorentz group at each point in the overlap region and we write:

$$AB^{-1} = L \tag{5.110}$$

where L is a Lorentz rotation matrix. We thus see that A and B can differ only by a Lorentz rotation at each point of the overlapping region,

$$A = LB \tag{5.111}$$

However, from the beginning the matrix A is arbitrary up to a constant Lorentz rotation, so we can just as well absorb an appropriate L into the matrix A to give

$$A = B \tag{5.112}$$

We then have one transformation which puts the metric in the Lorentz form in the *combined* neighborhoods of P_0 and P_1. This process can then be continued to any point nearby P_0 or P_1 just as one analytically continues a function in complex analysis. Eventually, any point of space can be included, so we have indeed extended our analysis to all space and shown that the entire Riemann space is pseudo-Euclidean.

Combined with the results of Sec. 5.2 and 5.5, the result of this section allows us to construct in summary the following list of *equivalent statements* about a given Riemann space.

1. The space is flat; that is, $R_{\alpha\beta\gamma\delta} = 0$.
2. The space is integrable, and parallel displacement is path-independent.
3. The space is pseudo-Euclidean, so there exists a coordinate system where the metric is everywhere constant.

Furthermore, for physical reasons, we shall restrict ourselves to considering the special case of a pseudo-Euclidean space in which the metric has signature $(1, -1, -1, -1)$.

5.7 The Einstein Field Equations for Free Space

We wish to obtain in this section an acceptable set of differential equations in tensor form to describe the gravitational field in free space; these equations must satisfy the four criteria which we stated in Sec. 5.1. The developments of the preceding sections lead us to expect that these equations should in some way involve the Riemann tensor $R_{\alpha\beta\gamma\delta}$ since this tensor appears to contain a great deal of information about the geometric structure of space. Note, for instance, the role played by the Riemann tensor in the parallel displacement of a vector in Eq. (5.82). Furthermore, we already know from the results of Sec. 5.2 that the special case of a gravity-free space with a Lorentz metric is correctly described by the equation $R_{\alpha\beta\gamma\delta} = 0$. Thus we expect the complete field equations to be some generalization of $R_{\alpha\beta\gamma\delta} = 0$, which, as is demanded by the third criterion of Sec. 5.1, still admits the Lorentz metric as one solution. In short, we wish in some way to *weaken* the above flat-space equation to admit more general solutions.

To obtain a clue as to how we might appropriately weaken this equation, let us consider Laplace's equation for the classical gravitational potential φ:

$$\sum_{i=1}^{3} \frac{\partial^2 \varphi}{\partial x^{i^2}} = \sum_{i=1}^{3} \varphi_{|i|i} = 0 \tag{5.113}$$

From Sec. 4.3, we know that, if Newton's second law of motion and the geodesic equation of motion are to yield approximately the same trajectories for slowly moving particles in a weak gravitational field, the g_{00} component of the metric tensor must be approximately given by

$$g_{00} = 1 + \frac{2\varphi}{c^2} \tag{5.114}$$

Therefore φ is given by

$$\varphi = \frac{c^2}{2}(g_{00} - 1) \tag{5.115}$$

and Laplace's equation may be written in terms of g_{00} as

$$\sum_{i=1}^{3} g_{00|i|i} = 0 \tag{5.116}$$

This equation, which must be an approximate form of the relativistic field equations (Sec. 7.2), involves second derivatives of the metric tensor with a summation over the repeated index i. In a covariant tensor equation the analogue of such a summation is a contraction, so we are led to expect a contraction to occur in the relativistic field equations. This observation suggests that we try weakening the equation $R_{\alpha\beta\gamma\delta} = 0$ by a contraction of the Riemann tensor. Fortunately, there is only one meaningful contraction of the Riemann tensor $R_{\alpha\beta\gamma\delta}$. Observe that a contraction between α and β or between γ and δ yields a null tensor since $R_{\alpha\beta\gamma\delta}$ is antisymmetric in these index pairs. Similarly, we see that contractions between α and γ, between α and δ, and between β and δ differ only in sign. Thus the only meaningful contraction which we may perform on $R_{\alpha\beta\gamma\delta} = 0$ yields the equation

$$R^{\alpha}{}_{\beta\alpha\delta} = R_{\beta\delta} = 0 \tag{5.117}$$

where $R_{\beta\delta}$ is termed the *contracted Riemann tensor* or *Ricci tensor*. Following Einstein, this is the equation we shall adopt to describe the gravitational field in free space.

Let us observe that the Ricci tensor is symmetric by virtue of the last symmetry relation in (5.51) for the full Riemann tensor. Indeed, using the symmetry of $g^{\alpha\beta}$, we obtain the chain of equations

$$R_{\eta\gamma} = R^{\alpha}{}_{\eta\alpha\gamma} = g^{\alpha\beta}R_{\beta\eta\alpha\gamma} = g^{\alpha\beta}R_{\alpha\gamma\beta\eta} = R_{\gamma\eta} \tag{5.117$'$}$$

Thus the Ricci tensor has 10 independent components.

Equation (5.117) clearly satisfies criteria (*a*) and (*b*) of Sec. 5.1: it is a tensor equation and was explicitly constructed so as to have the Lorentz metric as one solution. That it is second-order and quasi-linear in the components of $g_{\mu\nu}$, and thereby satisfies criteria (*b*) and (*d*), can be seen by writing out $R_{\beta\delta}$ in terms of the metric tensor in the form

$$\begin{aligned} R_{\beta\delta} &= \begin{Bmatrix} \alpha \\ \beta\ \alpha \end{Bmatrix}_{|\delta} - \begin{Bmatrix} \alpha \\ \beta\ \delta \end{Bmatrix}_{|\alpha} + \begin{Bmatrix} \alpha \\ \tau\ \delta \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\ \alpha \end{Bmatrix} - \begin{Bmatrix} \alpha \\ \tau\ \alpha \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\ \delta \end{Bmatrix} \\ &= \tfrac{1}{2}g^{\alpha\lambda}(g_{\beta\lambda|\alpha} + g_{\lambda\alpha|\beta} - g_{\beta\alpha|\lambda})_{|\delta} - \tfrac{1}{2}g^{\alpha\lambda}(g_{\beta\lambda|\delta} + g_{\lambda\delta|\beta} - g_{\beta\delta|\lambda})_{|\alpha} \\ &\qquad + (\text{terms involving } \textit{first} \text{ derivatives of } g_{\mu\nu}) \end{aligned} \tag{5.118}$$

Thus the four criteria of Sec. 5.1 are indeed satisfied by (5.117). Let us repeat in summary that the Einstein free-space field equations which we shall use in the following chapters are

$$R_{\beta\delta} = \begin{Bmatrix} \alpha \\ \beta\ \alpha \end{Bmatrix}_{|\delta} - \begin{Bmatrix} \alpha \\ \beta\ \delta \end{Bmatrix}_{|\alpha} + \begin{Bmatrix} \alpha \\ \tau\ \delta \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\ \alpha \end{Bmatrix} - \begin{Bmatrix} \alpha \\ \tau\ \alpha \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\ \delta \end{Bmatrix} = 0 \tag{5.119}$$

In Chap. 10 we shall also consider more general field equations for the interior of a distribution of matter and for space on a cosmic scale. Before doing this, however, we shall investigate the above system, which is the most important case.

5.8 The Divergenceless Form of the Einstein Field Equations

In special relativity we usually associate a vector or tensor of zero four-divergence with some conserved quantity; for instance, the zero divergence of the electromagnetic four-current j^{μ} is directly associated with the conservation of electric charge. It is thus useful in many branches of physics to write as many equations as possible in terms of divergenceless vectors or tensors. Equations written in this form deal directly with *persistent phenomena* instead of *transient events;* such a situation is certainly desirable whenever possible. We shall see in this section that it is possible to express the free-space gravitational field equations (5.119) quite simply in terms of a single divergenceless tensor. Let us begin by obtaining the divergence of $R_{\beta\delta}$. Raising the first two indices in the Bianchi identities (5.62) gives

$$\{R^{\alpha\eta}{}_{\beta\gamma\|\delta}\}_{(\beta,\gamma,\delta)} = R^{\alpha\eta}{}_{\beta\gamma\|\delta} + R^{\alpha\eta}{}_{\gamma\delta\|\beta} + R^{\alpha\eta}{}_{\delta\beta\|\gamma} = 0 \tag{5.120}$$

Contracting α with β and η with γ gives

$$R^{\alpha\eta}{}_{\alpha\eta\|\delta} + R^{\alpha\eta}{}_{\eta\delta\|\alpha} + R^{\alpha\eta}{}_{\delta\alpha\|\eta} = 0 \tag{5.121}$$

By definition of the Ricci tensor and by the symmetry properties of the Riemann tensor (5.51), we find

$$R^{\eta}{}_{\eta\|\delta} - R^{\alpha}{}_{\delta\|\alpha} - R^{\eta}{}_{\delta\|\eta} = 0 \tag{5.122}$$

Relabeling indices and rearranging terms, we then obtain

$$R^{\eta}{}_{\eta\|\delta} = 2R^{\beta}{}_{\delta\|\beta} \tag{5.123}$$

Denoting the doubly contracted Riemann tensor $R^{\eta}{}_{\eta}$ by R, the *Riemann scalar,* we can write this divergence in the form

$$\tfrac{1}{2}R_{\|\delta} = \tfrac{1}{2}g^{\beta}{}_{\delta}R_{\|\beta} = R^{\beta}{}_{\delta\|\beta} \tag{5.124}$$

For both indices in contravariant position, we have, then,

$$R^{\beta\delta}{}_{\|\beta} = \tfrac{1}{2}(g^{\beta\delta}R)_{\|\beta} \tag{5.125}$$

since the metric tensor has a zero covariant derivative. Thus the *Einstein tensor,* which we define as

$$G^{\beta\delta} = R^{\beta\delta} - \tfrac{1}{2}g^{\beta\delta}R \tag{5.126}$$

has zero divergence:

$$G^{\beta\delta}{}_{\|\beta} = 0 \tag{5.127}$$

Suppose, now, that the Riemann curvature tensor satisfies the free-space equations $R^{\beta\delta} = 0$. Then $R = 0$ also, and

$$G^{\beta\delta} = R^{\beta\delta} - \tfrac{1}{2}g^{\beta\delta}R = 0 \tag{5.128}$$

and the Einstein tensor is also zero. Conversely, if $G^{\beta\delta}$ is zero, then

$$G^{\beta}{}_{\delta} = R^{\beta}{}_{\delta} - \tfrac{1}{2}g^{\beta}{}_{\delta}R = 0 \tag{5.129}$$

Contracting this we see that R, the Riemann scalar, is zero:

$$G^{\beta}{}_{\beta} = 0 = R - \tfrac{1}{2}g^{\beta}{}_{\beta}R = R - 2R \tag{5.130}$$

Thus the Ricci tensor is also zero:

$$R^{\beta}{}_{\delta} = G^{\beta}{}_{\delta} + \tfrac{1}{2}g^{\beta}{}_{\delta}R = 0 \tag{5.131}$$

We conclude that $G^{\beta}{}_{\delta}$ *is zero if and only if* $R^{\beta}{}_{\delta}$ *is zero.* This allows us to write the Einstein field equations entirely in terms of the zero-divergence Einstein tensor:

$$G^{\beta}{}_{\delta} = R^{\beta}{}_{\delta} - \tfrac{1}{2}g^{\beta}{}_{\delta}R = 0 \tag{5.132}$$

This form of the equations will be extremely useful in the mathematical investigations of Chap. 8 and again in the physical developments of Chap. 10. Also in Chap. 10 the nature of the conservation law associated with the zero-divergence property of the Einstein tensor will become apparent.

5.9 The Riemann Tensor and Fields of Geodesics

We already know the role of geodesics in relativistic mechanics and in the theory of light rays. Hence the significance of the Riemann tensor in physical applications will be well illustrated by a formula which relates fields of geodesic curves in a Riemann space to the theory of the Riemann tensor. We consider a one-parameter family of geodesics $\Gamma(v)$

which are described by the system of equations

$$x^\alpha = x^\alpha(u,v) \tag{5.133}$$

where we suppose x^α to be twice continuously differentiable functions of u and v. The parameter v distinguishes between the different geodesics of the family, while the parameter u is the curve parameter on each $\Gamma(v)$. We have, for fixed v, the differential equation of the geodesics

$$\frac{\partial^2 x^\alpha}{\partial u^2} = -\left\{ {\alpha \atop \beta\ \gamma} \right\} \frac{\partial x^\beta}{\partial u} \frac{\partial x^\gamma}{\partial u} \qquad x^\alpha = x^\alpha(u,v) \tag{5.134}$$

We might in general identify u with the arc length on $\Gamma(v)$; however, we prefer to leave u to be defined just by (5.134) so that our reasoning remains valid also for null geodesics.

The family of geodesics gives rise to the field of tangent vectors

$$t^\alpha(u,v) = \frac{\partial x^\alpha(u,v)}{\partial u} \tag{5.135}$$

Let us also introduce the vector field

$$w^\alpha(u,v) = \frac{\partial x^\alpha(u,v)}{\partial v} \tag{5.136}$$

which describes the deviation of two points on two infinitesimally near geodesics which have the same parameter value u. We call w^α the vector of geodesic deviation in the geodesic field. The law of interchange of partial differentiation leads to the identity

$$\frac{\partial t^\alpha(u,v)}{\partial v} = \frac{\partial^2 x^\alpha}{\partial u\, \partial v} = \frac{\partial w^\alpha(u,v)}{\partial u} \tag{5.137}$$

We wish now to calculate the absolute derivative of the vector field $w^\alpha(u,v)$ on the geodesic $\Gamma(v)$. Let us use the definition (3.34) of this derivative and find by use of the Christoffel symbols instead of connections Γ

$$\begin{aligned} \frac{Dw^\alpha}{Du} &= \frac{\partial w^\alpha}{\partial u} + \left\{ {\alpha \atop \beta\ \gamma} \right\} \frac{\partial x^\beta}{\partial u} w^\gamma \\ &= \frac{\partial t^\alpha}{\partial v} + \left\{ {\alpha \atop \beta\ \gamma} \right\} t^\beta w^\gamma \end{aligned} \tag{5.138}$$

We have thus created a new vector field along each $\Gamma(v)$ and can therefore repeat the process of absolute differentiation. The remarkable fact appears that this second differentiation leads us directly to the Riemann curvature tensor. Indeed, we find

$$\frac{D^2 w^\alpha}{Du^2} = \frac{\partial}{\partial u}\left(\frac{Dw^\alpha}{Du}\right) + \left\{ {\alpha \atop \beta\ \gamma} \right\} t^\beta \frac{Dw^\gamma}{Du} \tag{5.139}$$

Inserting for Dw^α/Du from (5.138), we obtain by use of (5.135) and (5.137)

$$\begin{aligned} \frac{D^2 w^\alpha}{Du^2} = \frac{\partial}{\partial v}\left(\frac{\partial^2 x^\alpha}{\partial u^2}\right) &+ \left\{ {\alpha \atop \beta\ \gamma} \right\}_{|\delta} \frac{\partial x^\beta}{\partial u} \frac{\partial x^\gamma}{\partial v} \frac{\partial x^\delta}{\partial u} + \left\{ {\alpha \atop \beta\ \gamma} \right\} \frac{\partial^2 x^\beta}{\partial u^2} \frac{\partial x^\gamma}{\partial v} \\ &+ 2 \left\{ {\alpha \atop \beta\ \gamma} \right\} \frac{\partial x^\beta}{\partial u} \frac{\partial^2 x^\gamma}{\partial u\, \partial v} + \left\{ {\alpha \atop \beta\ \gamma} \right\} \left\{ {\gamma \atop \mu\ \nu} \right\} \frac{\partial x^\beta}{\partial u} \frac{\partial x^\mu}{\partial u} \frac{\partial x^\nu}{\partial v} \end{aligned} \tag{5.140}$$

We now apply the equation of the geodesics (5.134) in order to eliminate the terms $\partial^2 x^\alpha/\partial u^2$. A simple rearrangement and some obvious cancellations lead to the result

$$\begin{aligned} \frac{D^2 w^\alpha}{Du^2} = \Bigg(\left\{ {\alpha \atop \beta\ \delta} \right\}_{|\gamma} &- \left\{ {\alpha \atop \beta\ \gamma} \right\}_{|\delta} + \left\{ {\alpha \atop \gamma\ \tau} \right\} \left\{ {\tau \atop \beta\ \delta} \right\} \\ &- \left\{ {\alpha \atop \tau\ \delta} \right\} \left\{ {\tau \atop \beta\ \gamma} \right\} \Bigg) \frac{\partial x^\beta}{\partial u} \frac{\partial x^\gamma}{\partial u} w^\delta \end{aligned} \tag{5.141}$$

Comparing this result with the definition (5.10) of the Riemann curvature tensor, we can simplify (5.141) to

$$\frac{D^2 w^\alpha}{Du^2} = R^\alpha{}_{\beta\delta\gamma} \frac{\partial x^\beta}{\partial u} w^\delta \frac{\partial x^\gamma}{\partial u} \tag{5.142}$$

The equation system (5.142) is an ordinary second-order differential system for the geodesic deviation $w^\alpha(u,v)$ along a fixed geodesic $\Gamma(v)$. To illustrate the result, let us consider the case of Euclidean space in which the Riemann tensor vanishes identically. Hence the differential system (5.142) reduces to

$$\frac{D^2 w^\alpha}{Du^2} = 0 \tag{5.143}$$

We may choose a coordinate system in the large in which the Christoffel symbols vanish identically and use as curve parameter u the arc length s. We find, then,

$$\frac{d^2w^\alpha}{ds^2} = 0 \qquad w^\alpha(s) = a^\alpha s + b^\alpha \tag{5.144}$$

where a^α and b^α are constant vectors. Since the geodesics in Euclidean geometry are straight lines, we find the elementary result that the distance between two points moving with speed 1 along two given lines which are infinitesimally near is a linear function of time. The Riemann tensor measures, by (5.142), the departure from this linear behavior.

To illustrate the physical meaning of (5.142), let us consider an observer moving on a timelike geodesic and observing an object which moves near him on its own geodesic. The observer may use his arc length $ds = c\,dt = dx^0$ as time measure and will interpret the geodesic deviation $z^\alpha = \epsilon w^\alpha(s)$ as the Euclidean distance vector of the object. Here ϵ is a small positive factor measuring the distance of the object at the first moment of observation. According to (5.142), an acceleration of the object relative to the observer will be seen as

$$\frac{d^2z^\alpha}{dt^2} = R^\alpha{}_{0\delta 0} z^\delta \tag{5.145}$$

since in his coordinate system $t^\alpha \equiv (1,0,0,0)$.

Consider, on the other hand, the following problem in classical mechanics and Euclidean geometry. Suppose that observer and object move in a field of force which is mass-proportional and varies in space. If $F^i(x^k)$ is the vector of acceleration connected with this field, the object will be accelerated relative to the observer according to

$$\frac{d^2z^i}{dt^2} = F^i(x^k + z^k) - F^i(x^k) = \frac{\partial F^i}{\partial x^k} z^k + O(z^k) \tag{5.146}$$

if z^i is the vector from observer to object. The analogy between (5.145) and (5.146) is evident. We are led to the intuitive interpretation

$$R^i{}_{0k0} \leftrightarrow \frac{\partial F^i}{\partial x^k} \tag{5.147}$$

If the force field possesses a potential $\varphi(x^i)$ such that

$$F^i = -F_i = -\frac{\partial \varphi}{\partial x^i} \tag{5.148}$$

we find the correspondence

$$R^i{}_{0k0} \leftrightarrow -\frac{\partial^2\varphi}{\partial x^i\,\partial x^k} \tag{5.149}$$

Thus the condition for a Laplacian potential $\nabla^2\varphi = 0$ leads to

$$R^i{}_{0i0} = 0 \tag{5.150}$$

We come automatically to the Ricci tensor. To obtain an equation which is in tensor form and coordinate-invariant, we have to demand more generally that

$$R^\mu{}_{\alpha\mu\beta} = R_{\alpha\beta} = 0 \tag{5.151}$$

as a generalization of Laplace's equation. This consideration gives additional motivation to the choice of the field equations (5.117) to describe the gravitational field in empty space. The reader should also observe the analogy and differences between our present heuristic considerations and those of Sec. 5.7. While in the preceding section we considered the case of weak gravitational fields in the large, we dealt in this section with the approximate form of (5.117) for arbitrary gravitational fields, but in small distances. In both cases the analogy between the classical and relativistic formulas is striking.

To understand the physical significance of the equation of geodesic deviation (5.142) let us return to the Einstein box, discussed in the Introduction. We have seen that an observer in such a box could not decide whether the box was in a gravitational field or in accelerated motion in flat space by performing experiments with a single test body. This is due to the mathematical fact that at a given point in a Riemann space we can introduce geodesic coordinates; in this case the equation of motion of a freely falling body becomes $\ddot{x}^\alpha = 0$, which is precisely the same as the equation of force free motion in flat space, i.e. in special relativity. However, every laboratory has a finite size, and the Riemann tensor cannot be transformed away, so that we can, in principle, measure the inhomogeneities of the gravitational field by observing the *relative* motion of several freely falling bodies, using (5.142) or its classical analogue (5.146). For example, consider an experiment in an earth based laboratory in which two balls separated horizontally by 1 m are allowed to fall simultaneously a distance of 1 m. The two trajectories will converge by about 10^{-5} cm since the balls move toward the center of the earth. (To a first approximation they move on parallel trajectories, as they would in an accelerated laboratory in flat space.) Thus one determines that one is in a gravitational field and not in an accelerated laboratory. We emphasize that the measurement of a gravitational field, as manifested by a nonzero Riemann tensor, requires the observation of the trajectory of more than one freely falling body.

5.10 Algebraic Properties of the Riemann Tensor

The number of algebraically independent components of the Riemann tensor is reduced by symmetry; in dealing with solutions to the field equations the number of algebraically independent components is reduced yet further. In order to clarify this, Petrov has introduced a notation with indices that run over six values. With this notation we can also obtain an intrinsic classification of space-time geometries in terms of the algebraic properties of the Riemann tensor; this classification is analogous to the classification of electromagnetic fields as radiation and nonradiation fields.

The first set of two indices of the Riemann tensor $R_{\alpha\beta\gamma\delta}$ assumes 16 values: 00, 01, . . . , 33. However, from (5.51) the tensor is antisymmetric in these two indices, and so it is clear that we need the values of the tensor for only six pairs of indices. The same is true of the second set of two indices. We are therefore led to introduce the Petrov mapping which associates pairs of tensor indices with a single index as follows:

$$\begin{array}{lrcl} & \text{Tensor indices: } \alpha\beta & = & 23,\ 31,\ 12,\ 10,\ 20,\ 30;\ R_{\alpha\beta\gamma\delta} \\ (5.152) & \updownarrow & & \hspace{9em}\updownarrow \\ & \text{Petrov index*: } A & = & 1,\ \ 2,\ \ 3,\ \ 4,\ \ 5,\ \ 6;\ \mathbf{R}_{AB} \end{array}$$

The Riemann tensor is thus completely described by the 6×6 matrix $\mathbf{R}_{AB}$; all nonzero algebraically independent components of the tensor occur in the matrix $\mathbf{R}_{AB}$. Moreover, the symmetries in the index pairs expressed by Eqs. (5.51) are now embodied in the very simple statement that $\mathbf{R}_{AB}$ is symmetric, $\mathbf{R}_{AB} = \mathbf{R}_{BA}$. We next write $\mathbf{R}_{AB}$ in terms of 3×3 submatrices, two of which are symmetric by virtue of the symmetry of $\mathbf{R}_{AB}$:

$$(5.153) \qquad \mathbf{R}_{AB} = \begin{pmatrix} M & N \\ N^T & Q \end{pmatrix} \qquad M = M^T \qquad Q = Q^T$$

Then the cyclic symmetry (5.53) may conveniently be written as the condition that the trace of N vanish,

* We denote the Riemann tensor in the Petrov indication by $\mathbf{R}$ in order to avoid confusion with the contracted Riemann tensor. Observe that, for example, $\mathbf{R}_{11} = R_{2323}$, but $R_{11} = R^{\nu}{}_{1\nu1}$.

$$(5.154) \qquad \mathrm{Tr}(N) = 0$$

Thus the algebraic symmetries of the Riemann tensor are simply summarized by the statement that $\mathbf{R}_{AB}$ is a symmetric 6×6 matrix for which $\mathrm{Tr}(N) = 0$. From this we see in a very transparent way that it has 20 algebraically independent components.

If we work in a special coordinate system where the metric is Lorentzian at a given point, the field equations (5.117) yield further simplifications. For this purpose it is convenient to consider the mixed tensor $R^{\alpha\beta}{}_{\gamma\delta} \leftrightarrow \mathbf{R}^A{}_B$. To raise the Petrov index A we note from (5.152) that values of A from 1 to 3 correspond to two spatial tensor indices and therefore imply no sign change when raised. Values of A from 4 to 6 correspond to one space and one time tensor index and imply a sign change. Thus just as with tensor indices and the Lorentz metric we can raise a Petrov index by multiplication with a very simple matrix G^{AC}

$$(5.155) \qquad G^{AC} = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \qquad \mathbf{R}^A{}_B = G^{AC}\mathbf{R}_{CB} = \begin{pmatrix} M & N \\ -N^T & -Q \end{pmatrix}$$

Using the mixed Petrov matrix $\mathbf{R}^A{}_B$, we can write the field equations in elegant form

$$(5.156) \qquad \begin{aligned} R^0{}_0 &= 0 \Rightarrow \mathrm{Tr}(Q) = 0 \\ R^0{}_i &= 0 \Rightarrow N^T = N \\ R^i{}_j &= 0 \Rightarrow Q = -M \end{aligned}$$

If we combine these relations, implied by the field equations, with (5.153) and (5.154), implied by the algebraic symmetries, we see that $\mathbf{R}^A{}_B$ can be characterized by a simple statement

$$(5.157) \qquad \mathbf{R}^A{}_B = \begin{pmatrix} M & N \\ -N & M \end{pmatrix} \qquad M = M^T \quad N = N^T \quad \mathrm{Tr}(M) = \mathrm{Tr}(N) = 0$$

This matrix has only 10 algebraically independent components, but we emphasize that this will be true *only* in a coordinate system in which the metric is Lorentzian at the point of interest.

The mixed matrix (5.157) is not only very simple but is well suited to the study of invariants. This follows from the fact that a coordinate transformation of the Riemann tensor corresponds to a similarity transformation of the matrix $\mathbf{R}^A{}_B$. To show this we write the transformation of the tensor as

$$\bar{R}^{\alpha\beta}{}_{\gamma\delta} = \left(\frac{\partial \bar{x}^\alpha}{\partial x^\mu}\frac{\partial \bar{x}^\beta}{\partial x^\nu}\right) R^{\mu\nu}{}_{\lambda\tau} \left(\frac{\partial x^\lambda}{\partial \bar{x}^\gamma}\frac{\partial x^\tau}{\partial \bar{x}^\delta}\right) \tag{5.158}$$

Thanks to the symmetry properties of the tensor, we need not sum over all the indices $\mu\nu$ and $\lambda\tau$ but only over those corresponding to a Petrov index in (5.152). Thus we may write two equations that correspond term by term (see Exercise 5.7):

$$\bar{R}^{\alpha\beta}{}_{\gamma\delta} = \sum_{\left(\substack{\mu\nu,\ \gamma\delta \leftrightarrow \\ \text{Petrov}}\right)} \left(2\,\frac{\partial \bar{x}^\alpha}{\partial x^\mu}\frac{\partial \bar{x}^\beta}{\partial x^\nu}\right) R^{\mu\nu}{}_{\lambda\tau} \left(2\,\frac{\partial x^\lambda}{\partial \bar{x}^\gamma}\frac{\partial x^\tau}{\partial \bar{x}^\delta}\right) \tag{5.159}$$

$$\bar{\mathbf{R}}^A{}_B = S^A{}_C \mathbf{R}^C{}_D \tilde{S}^D{}_B$$

We easily verify that $S^A{}_C\tilde{S}^C{}_B = \delta^A{}_B$, that is, $\tilde{S}$ is the inverse matrix of S, and so we have established that $\bar{\mathbf{R}}^A{}_B$ and $\mathbf{R}^A{}_B$ are related by a similarity transformation. This result will be very useful since many algebraic properties of a matrix are invariant under similarity transformations.

To classify the space-time geometry, it is convenient to use not the Riemann tensor but its *self dual*, defined as

$$R^{(+)\alpha\beta}{}_{\gamma\delta} = R^{\alpha\beta}{}_{\gamma\delta} + i\,{}^*R^{\alpha\beta}{}_{\gamma\delta} \tag{5.160}$$

Here ${}^*R^{\alpha\beta}{}_{\gamma\delta}$ is the dual of the Riemann tensor (see Sec. 3.5)

$${}^*R_{\alpha\beta\gamma\delta} = \tfrac{1}{2}e_{\alpha\beta}{}^{\sigma\tau}R_{\sigma\tau\gamma\delta} \tag{5.161a}$$

$$R_{\alpha\beta\gamma\delta} = -\tfrac{1}{2}e_{\alpha\beta}{}^{\sigma\tau}\,{}^*R_{\sigma\tau\gamma\delta} \tag{5.161b}$$

We shall show that ${}^*R_{\alpha\beta\gamma\delta}$ has the same symmetry properties as $R_{\alpha\beta\gamma\delta}$ and also that the contracted tensor ${}^*R^\alpha{}_{\beta\alpha\delta}$ is zero; then the same reasoning used for $R^{\alpha\beta}{}_{\gamma\delta}$ earlier implies that ${}^*R^{\alpha\beta}{}_{\gamma\delta}$ corresponds to a 6×6 matrix ${}^*\mathbf{R}^A{}_B$, analogous to $\mathbf{R}^A{}_B$ in Eq. (5.157). To show first that the contracted tensor ${}^*R^\alpha{}_{\beta\alpha\delta}$ is zero we use the antisymmetry of $e^{\alpha\beta\sigma\tau}$ in the definition (5.161a) to write

$${}^*R^{\alpha\beta}{}_{\alpha\delta} = \tfrac{1}{2}e^{\alpha\beta\sigma\tau}R_{\sigma\tau\alpha\delta} = \tfrac{1}{2}e^{\alpha\beta\sigma\tau}\{R_{\sigma\tau\alpha\delta}\}_{(\sigma\tau\alpha)} \tag{5.162}$$

From the symmetry of the Riemann tensor expressed in (5.39) we see that this is indeed zero. To obtain the symmetries of the dual tensor we impose the field equations on (5.161b)

$$R^{\alpha\beta}{}_{\alpha\delta} = 0 = -\tfrac{1}{2}e^{\alpha\beta\sigma\tau}\,{}^*R_{\sigma\tau\alpha\delta} = -\tfrac{1}{2}e^{\alpha\beta\sigma\tau}\{{}^*R_{\sigma\tau\alpha\delta}\}_{(\sigma\tau\alpha)} \tag{5.163}$$

It is clear from this that $\{{}^*R_{\alpha\beta\gamma\delta}\}_{(\alpha\beta\gamma)}$ is zero, which implies that ${}^*R_{\alpha\beta\gamma\delta}$ has the same symmetries as $R_{\alpha\beta\gamma\delta}$; that is, Eq. (5.39) or equivalently (5.51) and (5.53) hold for ${}^*R_{\alpha\beta\gamma\delta}$. Thus ${}^*R^{\alpha\beta}{}_{\gamma\delta}$ corresponds to a 6×6 matrix ${}^*\mathbf{R}^A{}_B$ in the same way as the Riemann tensor corresponds to $\mathbf{R}^A{}_B$.

We continue to work with a local Lorentz metric, so that $e_{\alpha\beta\gamma\delta}$ is equal to $\epsilon_{\alpha\beta\gamma\delta}$. It is then easy to show from the definition (5.161) that the matrix ${}^*\mathbf{R}^A{}_B$ is given in terms of the 3×3 matrices M and N as

$${}^*\mathbf{R}^A{}_B = \begin{pmatrix} N & -M \\ M & N \end{pmatrix} \tag{5.164}$$

Thus corresponding to the self-dual tensor $R^{(+)\alpha\beta}{}_{\gamma\delta}$ is the matrix

$$\begin{gathered} \mathbf{R}^{(+)A}{}_B = \begin{pmatrix} P & -iP \\ iP & P \end{pmatrix} = P \otimes J \\ P \equiv M + iN \qquad J \equiv \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} \end{gathered} \tag{5.165}$$

The matrix P will be referred to as the *Petrov matrix*. Since it is a complex 3×3 traceless symmetric matrix, like $\mathbf{R}^A{}_B$ and $\mathbf{R}^{(+)A}{}_B$ it has 10 algebraically independent components.

Our classification will be according to the eigenvalues and the multiplicity of the eigenvectors of $R^{(+)\alpha\beta}{}_{\gamma\delta}$, the self-dual Riemann tensor, or equivalently the matrix $\mathbf{R}^{(+)A}{}_B$. These satisfy the equation

$$\begin{gathered} R^{(+)\alpha\beta}{}_{\gamma\delta}U^{\gamma\delta} = \lambda U^{\alpha\beta} \\ \updownarrow \\ \mathbf{R}^{(+)A}{}_B U^B = \lambda U^A \end{gathered} \tag{5.166}$$

and are therefore defined invariantly even though our analysis will use the special form of $\mathbf{R}^{(+)A}{}_B$ given in (5.162), which holds only in a special coordinate system. Specifically, $\mathbf{R}^{(+)A}{}_B$ transforms as in (5.159), U^B transforms via $\bar{U}^C = S^C{}_D U^D$, and λ is invariant.

The direct-product relation (5.165) between $\mathbf{R}^{(+)A}{}_B$ and the 3×3 Petrov matrix P allows us to express the nonzero eigenvalues and eigenvectors of $\mathbf{R}^{(+)A}{}_B$ in terms of those of P, in effect reducing the problem to that of a complex 3×3 matrix instead of a real 6×6. Indeed the eigenvectors of $\mathbf{R}^{(+)A}{}_B$ will be the direct product of those of P and J. We define for P and deduce for J the following eigenvector relations

$$P\zeta = \tau\zeta \qquad \zeta = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{5.167}$$

(5.167′)
$$J\eta = \sigma\eta \qquad \sigma = 0 \text{ and } \eta = \begin{pmatrix} 1 \\ -i \end{pmatrix} \qquad \text{or} \qquad \sigma = 2 \text{ and } \eta = \begin{pmatrix} 1 \\ i \end{pmatrix}$$

Thus $\mathbf{R}^{(+)A}{}_B$ has *at least* three zero eigenvalues, and so must the tensor $R^{(+)\alpha\beta}{}_{\gamma\delta}$. For our classification we consider only the remaining eigenvalues, which may or may not be zero; they and their corresponding eigenvectors are given explicitly by the direct products

(5.168)
$$\lambda = 2\tau \qquad U^A = \zeta \otimes \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ ia \\ ib \\ ic \end{pmatrix}$$

Moreover, since P is traceless by (5.157), the sum of the eigenvalues τ (and λ) must be zero, which leaves only two independent eigenvalues to consider.

To complete our classification of the interesting eigenvectors and eigenvalues of $\mathbf{R}^{(+)A}{}_B$ we consider the Jordan canonical form of P. According to Jordan, any $n \times n$ complex matrix is related by a similarity transformation to a matrix of the form

(5.169)
$$C = \begin{pmatrix} C_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & C_N \end{pmatrix} \qquad C_i = \begin{pmatrix} \tau_i & 1 & 0 & 0 & \cdots & \\ 0 & \tau_i & 1 & 0 & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ & & & & & \tau_i \end{pmatrix}$$

That is, the only nonzero elements of C_i are τ_i's along the diagonal and 1's along the first superdiagonal. (Observe that complex symmetric matrices cannot necessarily be put into diagonal form by a similarity transformation and that similarity transformations do not in general preserve symmetry.) The eigenvalues and eigenvectors of C are very easy to obtain; we easily verify that each C_i submatrix has eigenvalues equal to τ_i and only one eigenvector, (1,0, . . . ,0). Thus C has at most N distinct eigenvalues and N distinct eigenvectors. In general, the number N may be less than the dimension of the matrix, and so C may not have a full complement of eigenvectors. Since the eigenvalues and the algebraic relations among the eigenvectors are invariant properties under similarity transformations, the Jordan canonical form of P provides a complete description of these properties and will serve there-

fore to classify the space-time geometry. There are only five interesting possibilities for the Jordan canonical form of a traceless 3×3 matrix. We list these in Table 5.1, along with the consequent eigenvector and eigenvalue complement, according to Petrov's naming scheme; flat space, $\mathbf{R}^{(+)A}{}_B = 0$, is not included.

TABLE 5.1 PETROV CLASSIFICATION

Petrov type	Jordan form of P	Number of distinct eigenvectors	Number of distinct eigenvalues
I	$\begin{pmatrix} \tau_1 & 0 & 0 \\ 0 & \tau_2 & 0 \\ 0 & 0 & \tau_3 \end{pmatrix}$ $\tau_1 + \tau_2 + \tau_3 = 0$	3	3
ID	$\begin{pmatrix} \tau_1 & 0 & 0 \\ 0 & \tau_1 & 0 \\ 0 & 0 & -2\tau_1 \end{pmatrix}$	3	2
II	$\begin{pmatrix} \tau_1 & 1 & 0 \\ 0 & \tau_1 & 0 \\ 0 & 0 & -2\tau_1 \end{pmatrix}$	2	2
II*N*	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	2	All zero
III	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$	1	All zero

In this section we have made frequent use of a local Lorentz coördinate system for convenience, but we again emphasize that the invariant nature of the eigenvalue problem (5.166) gives our classification scheme invariant meaning. One use of the scheme is thereby immediately evident; two solutions of different type cannot be transformed into each other by any coordinate transformation. Such intrinsic differences are very important in the physical interpretations of general relativity.

We have found that unlike the electromagnetic field, which may be classified simply as radiation or nonradiation, the gravitational field has a much richer fivefold classification structure. This stems from the nonlinearity of the field equations; the lack of superposition in the gravitational theory prevents us from decomposing the field into simpler structures and therefore demands the richer classification scheme. In the problems for the following chapters we shall illustrate the use of the Petrov classification and its physical meaning. For example, the Schwarzschild solution, the gravitational analogue of the Coulomb field in electromagnetism, is Petrov type ID, whereas plane gravitational waves are type IIN.

Exercises

5.1 How many algebraically independent components does the Riemann tensor $R_{\alpha\beta\gamma\delta}$ have in two dimensions? How many does it have in n dimensions?

5.2 Show that for a metric of a two-dimensional space, of the form, $ds^2 = (dx^1)^2 + G^2(x^1)(dx^2)^2$ one has

$$R^1{}_{212} = -G\frac{d^2G}{(dx^1)^2}$$

Obtain all nonzero components of the Riemann tensor from this.

5.3 What is the Riemann tensor $R^\alpha{}_{\beta\gamma\eta}$ for the two-dimensional surface of a sphere? What is it for the surface of a cylinder? What is the Riemann scalar $R = R^\mu{}_\mu$ for these surfaces?

5.4 Consider a three-space imbedded in four-space in the particularly simple way $g_{00} = 1$, $g_{0i} = 0$, with the other g_{ij} independent of the time coordinate. How are the four-space Christoffel symbols related to the three-space symbols? What is the relation between $R_{\alpha\beta\gamma\delta}$ in four-space and R_{ijkl} in three-space?

5.5 (*continued*) What is R_{00}? What is G_{00}? What is the Riemann scalar in four-space? What is the Riemann scalar ${}^{(3)}R$ in three-space? Relate G_{00} and ${}^{(3)}R$.

5.6 Consider a geodesic triangle drawn on a sphere as follows. One vertex is at the north pole and the two others are on the equator, separated by 90°. Parallel-displace a vector around this triangle using the geometric result of Exercise 3.1. How is the vector changed after a complete circuit? (You may wish to choose a convenient initial orientation for simplicity.) Interpret your result and compare with Prob. 5.1.

5.7 Verify explicitly that (5.159) is equivalent to (5.158); i.e., the factors of 2 in (5.159) are correct.

5.8 Prove that the right dual tensor and the left dual tensor of the Riemann tensor, as defined in (5.160), are equal. (Is this true if the Einstein equations do not hold?)

Problems

5.1 In the plane, the total angle Δ through which the tangent to a closed curve turns in one circuit is always 2π. On a curved surface the corresponding Δ is defined as the sum of angles being measured in the successive local tangent planes. Δ will generally differ from 2π. A beautiful theorem (Gauss-Bonnet) says that

$$2\pi - \Delta = \int R\,dA$$

the integral of the curvature scalar over the enclosed area. Consider a sphere of radius a and on it a "geodesic triangle" formed by three geodesics making a right angle at each vertex (see Exercise 5.6). Test the theorem for both the areas that can be considered enclosed by this triangle.

5.2 A conformally flat space is defined as one with a metric tensor of the form $g_{\mu\nu} = f(x^\alpha)\eta_{\mu\nu}$, where f is an arbitrary positive function and $\eta_{\mu\nu}$ is the Lorentz metric. Show that for such a metric the *Weyl tensor*, defined as

$$C^\mu{}_{\nu\rho\sigma} = R^\mu{}_{\nu\rho\sigma} + g_{\nu\sigma}R^\mu{}_\rho - g_{\nu\rho}R^\mu{}_\sigma + R_{\nu\sigma}g^\mu{}_\rho - R_{\nu\rho}g^\mu{}_\sigma - \tfrac{1}{3}(g_{\nu\sigma}g^\mu{}_\rho - g_{\nu\rho}g^\mu{}_\sigma)R$$

is zero.

5.3 If the Einstein free-space field equations are satisfied, then the Weyl tensor is identical with the Riemann tensor. This implies that a conformally flat space for which $R_{\mu\nu} = 0$ is actually flat. Show this also by a coordinate transformation.

Bibliography

Standard texts on general relativity:

Bergmann, P. G. (1942): "An Introduction to the Theory of Relativity," New York.

Bergmann, P. G. (1962): The General Theory of Relativity, in Encyclopedia of Physics, vol. 4, Berlin-Göttingen-Heidelberg, pp. 203–272.

Eddington, A. S. (1924): "The Mathematical Theory of Relativity," 2d ed., London.

Einstein, A. (1956): "The Meaning of Relativity," 5th ed., Princeton, N.J.

Fock, V. (1959): "The Theory of Space, Time and Gravitation," New York.

Jordan, P. (1955): "Schwerkraft und Weltall," Brunswick.

Kretschmann, E. (1917): Über den physikalischen Sinn der Relativitätspostulate, *Ann. Physik*, **53** (4): 575–614.

Lichnerowicz, A. (1955): "Théories relativistes de la gravitation et de l'électromagnétisme," Paris.

Møller, C. (1952): "The Theory of Relativity," Oxford.

Pauli, W. (1958): "Theory of Relativity," New York.
Synge, J. L. (1960): "Relativity: The General Theory," Amsterdam.
Weyl, H. (1950): "Space, Time, Matter," New York.

Special topics:

Adler, R. J., and C. Sheffield (1972): Classification of Space-Times in General Relativity, *J. Math. Phys.*, **14**:465.
Barton, D., and J. P. Fitch (1971): General Relativity and the Application of Algebraic Manipulative Systems, *Commun. Am. Assoc. Comput. Mach.*, **14**:542.
Payne, A. B. (1972): Computation of the Einstein Tensor with Formac, *Computer Phys. Commun.*, **4**:100.
Petrov, A. Z. (1969): "Einstein Spaces," Oxford.

CHAPTER

6

The Schwarzschild Solution and Its Consequences: Experimental Tests of General Relativity

The free-space field equations (5.119) are nonlinear and hence difficult to solve. However, by imposing *symmetry conditions* dictated by physical arguments on the line element, we can greatly simplify the field equations in special cases. One such case is the *time-independent* and *spherically symmetric* line element; the resultant field equations were solved exactly by Schwarzschild in 1916. This solution is of particular importance since it corresponds to the basic one-body problem of classical astronomy. Indeed, the only reliable experimental verifications of the field equations (5.119), which we shall treat in Secs. 6.3, 6.5, and 6.6, are based on the Schwarzschild line element. In this chapter we shall obtain Schwarzschild's solution and discuss its consequences.

6.1 The Schwarzschild Solution

Consider the free-space field equations that we obtained in Chap. 5,

$$R_{\mu\nu} = \begin{Bmatrix} \beta \\ \beta\ \nu \end{Bmatrix}_{|\mu} - \begin{Bmatrix} \beta \\ \mu\ \nu \end{Bmatrix}_{|\beta} + \begin{Bmatrix} \beta \\ \tau\ \mu \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\ \nu \end{Bmatrix} - \begin{Bmatrix} \beta \\ \tau\ \beta \end{Bmatrix}\begin{Bmatrix} \tau \\ \mu\ \nu \end{Bmatrix} = 0 \tag{6.1}$$

We shall seek a solution which is time-independent and radially symmetric. By virtue of the requirement of radial symmetry, such a solution should represent the external field of a spherically symmetric body stationary at the origin. The limiting form of the line element at large distances from the origin may be expected to be Lorentzian and thus to be expressible in spherical coordinates r, θ, and φ as

$$ds^2 = c^2\,dt^2 - (dr^2 + r^2\,d\theta^2 + r^2 \sin^2\theta\,d\varphi^2) \qquad c\,dt = dx^0 \tag{6.2}$$

Next let us consider the above symmetry requirements and try to form the simplest line element which meets the demands of time-independence and radial symmetry. The reasoning which follows is based on plausibility only, in order to guess a heuristically reasonable and convenient line element. We should expect the line element to be invariant under inversion of the coordinate interval dx^0 (representing time); that is, ds^2 should be invariant under the replacement of dx^0 by $-dx^0$. This dictates that we use Gaussian coordinates in which the off-diagonal elements g_{0i} of the metric tensor are zero and the line element has the form $g_{00}\,(dx^0)^2 + g_{ik}\,dx^i\,dx^k$ with the g_{ik} independent of x^0. This is referred to as a *static* metric; it is to be distinguished from a metric which is merely independent of time, or *stationary*, as discussed in Sec. 3.7. Second, if there is to be *no preferred angular direction* in space, the line element should be independent of a change of $d\theta$ to $-d\theta$ and a change of $d\varphi$ to $-d\varphi$. This requires that there be no terms of the form $dr\,d\theta$, $d\theta\,d\varphi$, etc., in the line element, so the metric tensor must be entirely diagonal for the type of solution we desire. Thus we may write ds^2 as

$$ds^2 = Ac^2\,dt^2 - (B\,dr^2 + Cr^2\,d\theta^2 + Dr^2 \sin^2\theta\,d\varphi^2) \tag{6.3}$$

Furthermore, by our assumption of radial symmetry, the functions A, B, C, and D must be functions of r only. One more simplification of the form of the line element can be made on the basis of symmetry: we can suppose that the functions $C(r)$ and $D(r)$ which appear in (6.3) are equal. This can be seen as follows: A displacement by $\epsilon = r\,d\theta$ from the north pole ($\theta = 0$) corresponds to $ds^2 = -C\epsilon^2$, and a displacement by $\epsilon = r\,d\varphi$ along the equator ($\theta = \pi/2$) corresponds to $ds^2 = -D\epsilon^2$. If θ and φ are to represent angular coordinates, we should expect these quantities to be equal due to isotropy, which requires that $C \equiv D$. Then

$$ds^2 = Ac^2\,dt^2 - B\,dr^2 - C(r^2\,d\theta^2 + r^2 \sin^2\theta\,d\varphi^2) \tag{6.4}$$

The above line element represents the simplest form which is dictated by the symmetry requirements; however, it is possible to obtain a further simplification by a judicious choice of a radial coordinate. Specifically,

consider a radial coordinate defined by

$$\hat{r} = \sqrt{C(r)}\,r \tag{6.5}$$

It then follows that

$$Cr^2 = \hat{r}^2 \tag{6.6}$$

and

$$B\,dr^2 = \frac{B}{C}\left(1 + \frac{r}{2C}\frac{dC}{dr}\right)^{-2} d\hat{r}^2 \equiv \hat{B}\,d\hat{r}^2 \tag{6.7}$$

By means of (6.5) we can express $\hat{B}$ also as a function of the new radial coordinate $\hat{r}$. It is now clear that writing the line element (6.3) in terms of $\hat{r}$ by substituting from (6.6) and (6.7) yields a line element in which the coefficient of the angular term $d\theta^2 + \sin^2\theta\,d\varphi^2$ is 1. This, however, is equivalent to taking $C \equiv 1$ in the line element (6.4), so we conclude that, by a suitable choice of the radial coordinate, we can put the line element in the form

$$ds^2 = Ac^2\,dt^2 - B\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \tag{6.8}$$

with only two unknown functions of r. In order to exhibit clearly the signature of $g_{\mu\nu}$ and the sign of the determinant $\|g_{\mu\nu}\| = g$, let us write $A(r)$ as the intrinsically positive function $e^{\nu(r)}$ and $B(r)$ as $e^{\lambda(r)}$. The line element accordingly is written as

$$ds^2 = e^{\nu(r)}c^2\,dt^2 - e^{\lambda(r)}\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \tag{6.9}$$

This equation represents the final form of the line element we shall use in obtaining the Schwarzschild solution; as we have constructed it, the demands of time-independence and radial symmetry are clearly met.

The coordinate r used in (6.9) has a clear physical meaning. Consider a spherical surface defined by a constant value of r, on which points are labeled by θ and φ. The line element on this surface is

$$ds^2 = -r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \tag{6.10}$$

The physical length of the equator, defined as the line $\theta = \pi/2$, is obtained by integrating $\sqrt{-ds^2}$ from (6.10) from $\varphi = 0$ to 2π, which gives

$$l = \int_0^{2\pi} \sqrt{-ds^2} = \int_0^{2\pi} r\,d\varphi = 2\pi r \tag{6.11}$$

This is identical to the flat-space result for a spherical surface. Thus by a measurement of the physical length of a great circle we can determine the value of the coordinate r for the sphere considered. Similarly it is easy to show that the physical area of such a sphere is $4\pi r^2$, as in flat space, which again allows us to determine the value of r. It is thereby clear that r is geometrically distinguished, and, moreover, the three space coordinates r, θ, and φ correspond closely to the variables used by astronomers in actual observations.

Even with the simplified metric form of (6.9), the work of computing the 40 Christoffel symbols appearing in the field equations is rather tedious. There is, however, a simple and convenient artifice which we may use to obtain all the nonzero Christoffel symbols. The Euler-Lagrange equations of the geodesic lines in the form

$$\ddot{x}^\alpha + \begin{Bmatrix} \alpha \\ \beta\ \eta \end{Bmatrix} \dot{x}^\beta \dot{x}^\eta = 0 \qquad \dot{x}^r \equiv \frac{dx^r}{ds} \tag{6.12}$$

contain all the Christoffel symbols. Conversely, if we know the Euler-Lagrange equations for the geodesic lines, we can identify all the nonzero Christoffel symbols; the identification is especially simple for a diagonal metric tensor. The Euler-Lagrange equations can be obtained from the variational problem

$$\delta \int ds = \delta \int [e^\nu (\dot{x}^0)^2 - (e^\lambda \dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta\, \dot{\varphi}^2)]^{1/2}\, ds = 0 \tag{6.13}$$

However, since we are using s as the variable of integration, we can just as well consider the equivalent and somewhat simpler variational problem in which we square the integrand of (6.13), according to the results of Sec. 2.3. That is, instead of (6.13), we use

$$\delta \int [e^\nu (\dot{x}^0)^2 - (e^\lambda \dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta\, \dot{\varphi}^2)]\, ds = 0 \tag{6.14}$$

Using F to represent the integrand, we write the Euler-Lagrange equations for this variational problem in the form

$$\frac{d}{ds}\left(\frac{\partial F}{\partial \dot{x}^\alpha}\right) = \frac{\partial F}{\partial x^\alpha} \tag{6.15}$$

We shall write out the four equations (6.15) explicitly and compare with the form in (6.12). We identify $x^1 = r$, $x^2 = \theta$, and $x^3 = \varphi$. The comparison will allow us to write out explicitly the non-vanishing Christoffel symbols.

The Euler-Lagrange equation for $x^\alpha = x^0$ is obtained from (6.14) as

$$\frac{d}{ds}(2e^\nu \dot{x}^0) = 0 \tag{6.16}$$

Denoting differentiation with respect to r by a prime, we then have

$$\ddot{x}^0 + \nu' \dot{r} \dot{x}^0 = 0 \tag{6.17}$$

This is a particularly simple equation because of the time-independence of the line element. Comparing (6.12) with (6.17), we obtain the following nonzero Christoffel symbols whose upper indices are zero:

$$\begin{Bmatrix} 0 \\ 1\ 0 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0\ 1 \end{Bmatrix} = \tfrac{1}{2}\nu' \tag{6.18}$$

Similarly, for the variable r, we have the Euler-Lagrange equation

$$\ddot{r} + \tfrac{1}{2}\lambda' \dot{r}^2 + \tfrac{1}{2}\nu' e^{\nu-\lambda}(\dot{x}^0)^2 - e^{-\lambda} r \dot{\theta}^2 - r \sin^2\theta\, \dot{\varphi}^2 e^{-\lambda} = 0 \tag{6.19}$$

The only nonzero Christoffel symbols with the upper indices 1 are therefore

$$\begin{aligned} &\begin{Bmatrix} 1 \\ 0\ 0 \end{Bmatrix} = \tfrac{1}{2}\nu' e^{\nu-\lambda} && \begin{Bmatrix} 1 \\ 1\ 1 \end{Bmatrix} = \tfrac{1}{2}\lambda' \\ &\begin{Bmatrix} 1 \\ 2\ 2 \end{Bmatrix} = -e^{-\lambda} r && \begin{Bmatrix} 1 \\ 3\ 3 \end{Bmatrix} = -r\sin^2\theta\, e^{-\lambda} \end{aligned} \tag{6.20}$$

Continuing in this way, we obtain the Euler-Lagrange equation for the variable θ,

$$\ddot{\theta} + \frac{2}{r}\dot{\theta}\dot{r} - \sin\theta\cos\theta\, \dot{\varphi}^2 = 0 \tag{6.21}$$

so the corresponding nonzero Christoffel symbols are

$$\begin{Bmatrix} 2 \\ 2\ 1 \end{Bmatrix} = \begin{Bmatrix} 2 \\ 1\ 2 \end{Bmatrix} = \frac{1}{r} \qquad \begin{Bmatrix} 2 \\ 3\ 3 \end{Bmatrix} = -\sin\theta\cos\theta \tag{6.22}$$

The final Euler-Lagrange equation for the variable φ is

$$\ddot{\varphi} + 2\cot\theta\, \dot{\varphi}\dot{\theta} + \frac{2}{r}\dot{r}\dot{\varphi} = 0 \tag{6.23}$$

so the corresponding Christoffel symbols are

$$(6.24)\qquad \begin{Bmatrix} 3 \\ 2\;3 \end{Bmatrix} = \begin{Bmatrix} 3 \\ 3\;2 \end{Bmatrix} = \cot\theta \qquad \begin{Bmatrix} 3 \\ 1\;3 \end{Bmatrix} = \begin{Bmatrix} 3 \\ 3\;1 \end{Bmatrix} = \frac{1}{r}$$

We now have all the nonzero Christoffel symbols displayed in (6.18), (6.20), (6.22), and (6.24).

The field equations (6.1) contain contracted Christoffel symbols of the form $\begin{Bmatrix} \tau \\ \tau\;\beta \end{Bmatrix}$; these may be written in the convenient form $(\log\sqrt{-g})_{|\beta}$ according to (3.11), which allows us to write the field equations as

$$(6.25)\qquad R_{\mu\nu} = (\log\sqrt{-g})_{|\mu|\nu} - \begin{Bmatrix} \alpha \\ \mu\;\nu \end{Bmatrix}_{|\alpha} + \begin{Bmatrix} \beta \\ \tau\;\mu \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\;\nu \end{Bmatrix} - \begin{Bmatrix} \tau \\ \mu\;\nu \end{Bmatrix}(\log\sqrt{-g})_{|\tau} = 0$$

Using the line element (6.9), we can write the expression $\log\sqrt{-g}$ explicitly in terms of the coordinates r, θ, φ, and t. The metric tensor is

$$(6.26)\qquad g_{\mu\nu} = \begin{pmatrix} e^{\nu(r)} & 0 & 0 & 0 \\ 0 & -e^{\lambda(r)} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2\sin^2\theta \end{pmatrix}$$

so the determinant g is

$$(6.27)\qquad g = \|g_{\mu\nu}\| = -e^{\nu(r)+\lambda(r)}r^4\sin^2\theta$$

Thus

$$(6.28)\qquad \log\sqrt{-g} = \frac{\nu+\lambda}{2} + 2\log r + \log|\sin\theta|$$

We are now ready to write out the field equations (6.25) in terms of r, θ, φ, and t.

First let us consider the $\mu = \nu = 0$ component of (6.25)

$$(6.29)\qquad R_{00} = (\log\sqrt{-g})_{|0|0} - \begin{Bmatrix} \alpha \\ 0\;0 \end{Bmatrix}_{|\alpha} + \begin{Bmatrix} \beta \\ \tau\;0 \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\;0 \end{Bmatrix} - \begin{Bmatrix} \tau \\ 0\;0 \end{Bmatrix}(\log\sqrt{-g})_{|\tau} = 0$$

Many of the terms appearing in this equation are identically zero; the nonzero terms are displayed in (6.18), (6.20), (6.22), (6.24), and (6.28), and we are left with

$$(6.30)\qquad R_{00} = -\begin{Bmatrix} 1 \\ 0\;0 \end{Bmatrix}_{|1} + 2\begin{Bmatrix} 1 \\ 0\;0 \end{Bmatrix}\cdot\begin{Bmatrix} 0 \\ 0\;1 \end{Bmatrix} - \begin{Bmatrix} 1 \\ 0\;0 \end{Bmatrix}(\log\sqrt{-g})_{|1}$$

$$= -(\tfrac{1}{2}\nu' e^{\nu-\lambda})' + (\tfrac{1}{2}\nu'^2 e^{\nu-\lambda}) - (\tfrac{1}{2}\nu' e^{\nu-\lambda})\left(\frac{\nu'+\lambda'}{2} + \frac{2}{r}\right) = 0$$

This reduces to

$$R_{00} = \frac{-e^{\nu-\lambda}}{2}\left(\nu'' + \frac{\nu'^2}{2} - \frac{\lambda'\nu'}{2} + \frac{2\nu'}{r}\right)$$

$$(6.31)\qquad \nu'' + \tfrac{1}{2}\nu'^2 - \tfrac{1}{2}\lambda'\nu' + \frac{2\nu'}{r} = 0$$

We proceed in similar fashion for the $\mu = \nu = 1$ component of Eq. (6.25)

$$(6.32)\qquad R_{11} = (\log\sqrt{-g})_{|1|1} - \begin{Bmatrix} \alpha \\ 1\;1 \end{Bmatrix}_{|\alpha} + \begin{Bmatrix} \beta \\ \tau\;1 \end{Bmatrix}\begin{Bmatrix} \tau \\ \beta\;1 \end{Bmatrix} - \begin{Bmatrix} \tau \\ 1\;1 \end{Bmatrix}(\log\sqrt{-g})_{|\tau} = 0$$

Discarding identically vanishing terms, we obtain

$$(6.33)\qquad R_{11} = (\log\sqrt{-g})_{|1|1} - \begin{Bmatrix} 1 \\ 1\;1 \end{Bmatrix}_{|1} + \begin{Bmatrix} 0 \\ 1\;0 \end{Bmatrix}\begin{Bmatrix} 0 \\ 1\;0 \end{Bmatrix} + \begin{Bmatrix} 1 \\ 1\;1 \end{Bmatrix}\begin{Bmatrix} 1 \\ 1\;1 \end{Bmatrix}$$

$$+ \begin{Bmatrix} 2 \\ 2\;1 \end{Bmatrix}\begin{Bmatrix} 2 \\ 2\;1 \end{Bmatrix} + \begin{Bmatrix} 3 \\ 3\;1 \end{Bmatrix}\begin{Bmatrix} 3 \\ 3\;1 \end{Bmatrix} - \begin{Bmatrix} 1 \\ 1\;1 \end{Bmatrix}(\log\sqrt{-g})_{|1} = 0$$

that is,

$$(6.34)\qquad R_{11} = \left(\frac{\nu''+\lambda''}{2} - \frac{2}{r^2}\right) - \tfrac{1}{2}\lambda'' + \tfrac{1}{4}\nu'^2 + \tfrac{1}{4}\lambda'^2 + \frac{2}{r^2} - \tfrac{1}{2}\lambda'\left(\frac{\lambda'+\nu'}{2} + \frac{2}{r}\right) = 0$$

This reduces to

$$R_{11} = \tfrac{1}{2}\left(\nu'' + \tfrac{1}{2}\nu'^2 - \tfrac{1}{2}\lambda'\nu' - \frac{2\lambda'}{r}\right)$$

$$(6.35)\qquad \nu'' + \tfrac{1}{2}\nu'^2 - \tfrac{1}{2}\lambda'\nu' - \frac{2\lambda'}{r} = 0$$

Equations (6.31) and (6.35) now represent a system of two ordinary differential equations which we may solve for the functions $\nu(r)$ and $\lambda(r)$. Subtraction of (6.35) from (6.31) yields

$$\nu' + \lambda' = 0 \tag{6.36}$$

Thus

$$\nu + \lambda = \text{const} = k \tag{6.37}$$

We can choose the constant k to be zero by a simple device. Replace the time coordinate t by another coordinate $t \exp(k/2)$; from (6.9) it is clear that this is equivalent to replacing ν by $\nu + k$, so that (6.37) becomes

$$\lambda = -\nu \tag{6.38}$$

We shall see that this choice of time coordinate has the very desirable feature of making the line element asymptotically equal to the flat-space line element (6.2). The coordinate t which we select in this way will be seen to correspond to the physical time as measured by an observer at infinity.

Substituting $-\lambda$ for ν in (6.35), we obtain a second-order ordinary differential equation for $\lambda(r)$

$$\lambda'' - \lambda'^2 + \frac{2\lambda'}{r} = 0 \tag{6.39}$$

This can be more conveniently written as

$$(re^{-\lambda})'' = 0 \tag{6.40}$$

Integration is then trivial, and we have

$$(re^{-\lambda})' = \text{const} \tag{6.41}$$

It will be convenient to leave (6.41) as it stands (with an undetermined constant) and proceed to the component corresponding to R_{22} of the system of equations (6.25); the reason for this will be apparent in the next paragraph.

Proceeding as with the R_{00} and R_{11} equations, we have

$$R_{22} = (\log\sqrt{-g})_{|2|2} - \left\{ {\alpha \atop 2\ 2} \right\}_{|\alpha} + \left\{ {\beta \atop \tau\ 2} \right\}\left\{ {\tau \atop \beta\ 2} \right\} - \left\{ {\tau \atop 2\ 2} \right\}(\log\sqrt{-g})_{|\tau} \tag{6.42}$$

Substitution of the nonzero terms from (6.18), (6.20), (6.22), (6.24), and (6.28) gives

$$R_{22} = (\log\sqrt{-g})_{|2|2} - \left\{ {1 \atop 2\ 2} \right\}_{|1} + \left\{ {1 \atop 2\ 2} \right\}\left\{ {2 \atop 2\ 1} \right\} + \left\{ {2 \atop 2\ 1} \right\}\left\{ {1 \atop 2\ 2} \right\} + \left\{ {3 \atop 2\ 3} \right\}\left\{ {3 \atop 2\ 3} \right\} - \left\{ {1 \atop 2\ 2} \right\}(\log\sqrt{-g})_{|1} = 0 \tag{6.43}$$

that is,

$$R_{22} = \frac{\partial^2}{\partial\theta^2}(\log|\sin\theta|) + (e^{-\lambda}r)' + 2(-e^{-\lambda}) + \cot^2\theta + e^{-\lambda}r\left(\frac{\lambda' + \nu'}{2} + \frac{2}{r}\right) = 0 \tag{6.44}$$

By virtue of (6.36) this simplifies to

$$(e^{-\lambda}r)' = 1 \tag{6.45}$$

This is precisely the same as Eq. (6.41) except that the unknown constant which appeared in (6.41) is now identified as 1. Integration immediately gives

$$e^{-\lambda}r = r - 2m \tag{6.46}$$

where $-2m$ is an arbitrary constant of integration. Thus, from the three equations $R_{00} = R_{11} = R_{22} = 0$, we have solved for the functions $\nu(r)$ and $\lambda(r)$ which appear in the line element (6.9)

$$\begin{aligned} e^{\nu} &= e^{-\lambda} = 1 - \frac{2m}{r} \\ e^{\lambda} &= \frac{1}{1 - 2m/r} \end{aligned} \tag{6.47}$$

Consider for a moment the result (6.47). As we noted, it was only necessary to use three of the ten equations in the system (6.25) to obtain what appears to be a complete solution (6.47). Apparently, then, we have a consistency problem remaining: Are the other seven equations in the system (6.25) *consistent* with the solution (6.47)? We shall show that the remaining diagonal element R_{33} of the Ricci tensor is indeed zero by virtue of the solution (6.47), so that (6.47) and the equation $R_{33} = 0$

are consistent. As before, we write

$$(6.48)\quad R_{33} = (\log \sqrt{-g})_{|3|3} - \left\{ {\alpha \atop 3\ 3} \right\}_{|\alpha} + \left\{ {\beta \atop \tau\ 3} \right\} \left\{ {\tau \atop \beta\ 3} \right\} - \left\{ {\tau \atop 3\ 3} \right\} (\log \sqrt{-g})_{|\tau} = 0$$

Discarding the identically zero terms, we get

$$(6.49)\quad R_{33} = - \left\{ {1 \atop 3\ 3} \right\}_{|1} - \left\{ {2 \atop 3\ 3} \right\}_{|2} + 2 \left\{ {3 \atop 1\ 3} \right\} \left\{ {1 \atop 3\ 3} \right\} + 2 \left\{ {3 \atop 3\ 2} \right\} \left\{ {2 \atop 3\ 3} \right\} - \left\{ {1 \atop 3\ 3} \right\} (\log \sqrt{-g})_{|1} - \left\{ {2 \atop 3\ 3} \right\} (\log \sqrt{-g})_{|2} = 0$$

Observe that, by virtue of (6.28) and (6.38), we have

$$(6.50)\quad \log \sqrt{-g} = 2 \log r + \log |\sin \theta|$$

Hence

$$(6.51)\quad (re^{-\lambda} \sin^2 \theta)' + \frac{\partial}{\partial \theta} (\sin \theta \cos \theta) + 2(-e^{-\lambda} \sin^2 \theta - \cot \theta \sin \theta \cos \theta) + re^{-\lambda} \sin^2 \theta \left(\frac{2}{r}\right) + \cos^2 \theta = 0$$

This simplifies to

$$(6.52)\quad \sin^2 \theta[(e^{-\lambda}r)' - 1] = 0$$

By virtue of (6.45) this is identically zero, so we see that the equation $R_{33} = 0$ is indeed consistent with (6.47). We leave it to the reader to verify that all the off-diagonal elements of the contracted Riemann tensor are identically zero when explicitly written in terms of the coordinates and that (6.47) is therefore a completely consistent solution of (6.25).

Let us now summarize the results of this section by exhibiting the *Schwarzschild line element*

$$(6.53)\quad ds^2 = \left(1 - \frac{2m}{r}\right)(dx^0)^2 - \frac{dr^2}{1 - 2m/r} - r^2(d\theta^2 + \sin^2 \theta\, d\varphi^2)$$

This result must be considered to be the main achievement of general relativity theory in the field of celestial mechanics; it is an exact solution,

which corresponds historically to Newton's treatment of the $1/r^2$ force law of classical gravitational theory. The rest of this chapter will be devoted to investigating the physical consequences of the line element (6.53).

It is evident that the Schwarzschild line element approaches the flat-space form (6.2) at large r. We may therefore identify t with the time measured by an observer at a large distance from the origin. Thus the coordinate time is, in this sense, a distinguished coordinate. It is important to keep in mind that the physical meaning of the coordinates is intimately related to the metric, as this example demonstrates.

The unknown constant of integration m which appears in the Schwarzschild line element can be determined by an appeal to correspondence with Newtonian theory. Recall that, in Sec. 4.3, we found that a geometric theory of gravitation will reduce in the classical limit of weak fields and slowly moving bodies to the Newtonian theory if $g_{00} \cong 1 + 2\varphi/c^2$; φ is the classical potential for the gravitational field. In the present case of a point mass, φ is simply $-\kappa M/r$, where M is the mass of the particle and $\kappa = 6.67 \times 10^{-8}$ dyne-cm^2/g^2 is the gravitational constant. Thus, in the classical limit, $g_{00} \cong 1 - 2\kappa M/c^2r$. Comparing this with (6.53), we see that

$$(6.54)\quad m = \frac{\kappa M}{c^2}$$

The constant m has the units of distance and will be referred to as the *geometric mass* of the central body.

We have here obtained the Schwarzschild solution by imposing the conditions of spherical symmetry and time-independence. However, it can be proved (Birkhoff, 1923) that the requirement of time-independence is superfluous and that any spherically symmetric distribution of matter, even if in radial motion, leads to the *same* line element exterior to the matter distribution. This result is called *Birkhoff's theorem*. The derivation is straightforward but more cumbersome than that presented in the text, since λ and ν are treated as functions of r and t (see Prob. 6.1). A consequence of Birkhoff's theorem is that a radially pulsating distribution of matter can emit no gravitational waves since the metric exterior to the distribution is static. Such waves can therefore be emitted only by more complicated deformations of a massive body.

Before continuing to the next section, it should be noted that, on the spherical shell $r = 2m$, the coefficient of dr^2 in the Schwarzschild line element becomes infinite and the coefficient of $(dx^0)^2$ is zero; $r = 2m$ is called the *Schwarzschild radius*. For ordinary stars this is characteristically a very small number; for the sun the Schwarzschild radius is about

3 km, which is well *inside* the sun, where the free-space field equations (6.25) are *not* valid and the Schwarzschild line element is *not* an appropriate description of the space-time geometry. The existence of this singularity is therefore of no consequence for the description of planetary motion.

It is clear that a star of roughly solar mass would have to be compressed to exceedingly high density before the bulk of its mass could be inside the Schwarzschild radius. The study of just this situation has become of great interest in recent years since it now appears likely from theoretical studies that a significant fraction of the stars in the universe may actually reach and exceed such densities in the process of gravitational collapse, which occurs at the end of their existence as normal stars. We shall return to this subject later in this chapter and again in Chap. 14.

6.2 The Schwarzschild Solution in Isotropic Coordinates

In the preceding section we obtained the Schwarzschild solution (6.53) in terms of a set of spherical polar coordinates: r, θ, φ, and t. The choice of this particular set of coordinates was motivated by the radial symmetry, time-independence, and relative simplicity required of the basic line element (6.9). However, it is characteristic of general relativity that there are usually many convenient coordinate systems available in which to work, and the coordinates r, θ, φ, and t in which (6.53) is expressed are not the only coordinates which correspond to our intuitive notions of radial and angular markers. In this section we shall consider another convenient set of coordinates and investigate the Schwarzschild line element expressed in the new coordinates.

The main reason for seeking an alternative set of coordinates is that we would like to express ds^2 in a form which is independent of the particular *space* coordinates used. More specifically, we would like to put the line element in the form

$$ds^2 = A(r)(dx^0)^2 - B(r)\, d\sigma^2 \tag{6.55}$$

where $d\sigma^2$ is $dx^2 + dy^2 + dz^2$ in Cartesian coordinates or $dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2$ in spherical coordinates, etc. This sort of line element agrees most closely with our intuitive notion of space, which is based mainly on Euclidean geometry. Indeed, to illustrate this, consider two vectors in three dimensions, ξ^i and η^i. In the metric (6.55) the cosine of the angle between these vectors, $\xi^i\eta_i/|\xi|\,|\eta|$, is the same as if we were in Euclidean space; this is due to the fact that the factor $B(r)$ cancels in the above ratio.

For this reason the line element (6.55) is called a *conformal line element.* The coordinates in which the line element takes the form (6.55) are called *isotropic coordinates.*

To obtain isotropic coordinates we shall attempt to use the following particularly simple coordinate transformation: The coordinates θ, φ, and t remain unchanged, while a radial coordinate $\rho(r)$ replaces r. In terms of these coordinates we ask that the Schwarzschild line element have the isotropic form

$$\begin{aligned} ds^2 &= \left(1 - \frac{2m}{r}\right) c^2\, dt^2 - \lambda^2(\rho)[d\rho^2 + \rho^2(d\theta^2 + \sin^2\theta\, d\varphi^2)] \\ &= \left(1 - \frac{2m}{r}\right) c^2\, dt^2 - \lambda^2(\rho)\, d\sigma^2 \end{aligned} \tag{6.56}$$

These demands lead to the mathematical problem of finding two functions of ρ, $r(\rho)$ and $\lambda(\rho)$, for which the two forms of the Schwarzschild line element, (6.53) and (6.56), are consistent. Comparing the coefficients of the angular interval $(d\theta^2 + \sin^2\theta\, d\varphi^2)$ in (6.53) and (6.56), we see that we must have

$$r^2 = \lambda^2\rho^2 \tag{6.57}$$

A similar comparison of the radial intervals gives

$$\frac{dr^2}{1 - 2m/r} = \lambda^2\, d\rho^2 \tag{6.58}$$

Substituting for λ^2 from (6.57) and taking the square root, we obtain an ordinary differential equation for $r(\rho)$:

$$\frac{\pm dr}{\sqrt{r^2 - 2mr}} = \frac{d\rho}{\rho} \tag{6.59}$$

An easy integration then yields

$$\pm \log\,[(r^2 - 2mr)^{1/2} + (r - m)] = \log\rho + \text{const} \tag{6.60}$$

To evaluate the constant and determine the sign of the left side of (6.60), consider r much larger than $2m$; asymptotically we must have

$$\pm \log\,(2r) = \log\rho + \text{const} \tag{6.61}$$

For large radial distances we wish r and ρ to be asymptotically equal, so we must choose the plus sign and take the constant to be log 2. Equation (6.60) then gives

$$\sqrt{r^2 - 2mr} + (r - m) = 2\rho \tag{6.62}$$

In order to solve this algebraic equation for r as a function of ρ, note that

$$[(r - m) + \sqrt{r^2 - 2mr}][(r - m) - \sqrt{r^2 - 2mr}] = m^2 \tag{6.63}$$

Dividing this by (6.62), we obtain

$$(r - m) - \sqrt{r^2 - 2mr} = \frac{m^2}{2\rho} \tag{6.64}$$

Addition of (6.64) above to (6.62) yields

$$r - m = \rho + \frac{m^2}{4\rho} \tag{6.65}$$

Thus, finally,

$$r = \rho + \frac{m^2}{4\rho} + m = \rho\left(1 + \frac{m}{2\rho}\right)^2 \tag{6.66}$$

From Eq. (6.57) and the above it follows that the function $\lambda(\rho)$ is

$$\lambda = \frac{r}{\rho} = \left(1 + \frac{m}{2\rho}\right)^2 \tag{6.67}$$

Let us now return to the isotropic form of the Schwarzschild line element (6.56) and express it in terms of the coordinate ρ. According to (6.66), the coefficient of dt is

$$\left(1 - \frac{2m}{r}\right) = 1 - \frac{2m}{\rho(1 + m/2\rho)^2} = \frac{(1 - m/2\rho)^2}{(1 + m/2\rho)^2} \tag{6.68}$$

Thus, by (6.56), (6.67), and (6.68), the Schwarzschild line element in terms of *isotropic* coordinates is

$$\begin{aligned} ds^2 &= \frac{(1 - m/2\rho)^2}{(1 + m/2\rho)^2} c^2\, dt^2 \\ &\quad - \left(1 + \frac{m}{2\rho}\right)^4 (d\rho^2 + \rho^2\, d\theta^2 + \rho^2 \sin^2\theta\, d\varphi^2) \\ &= \frac{(1 - m/2\rho)^2}{(1 + m/2\rho)^2} c^2\, dt^2 - \left(1 + \frac{m}{2\rho}\right)^4 d\sigma^2 \end{aligned} \tag{6.69}$$

We have now succeeded in putting the Schwarzschild line element in isotropic form; in such a form it should still be directly comparable with the approximate solution (4.142) obtained by a correspondence argument. In order to make a comparison, note that the constant of integration m which appears in the isotropic line element (6.69) evidently serves as a measure of how much the line element differs from the Lorentzian form $c^2\, dt^2 - d\sigma^2$; indeed, $m = 0$ gives precisely the Lorentzian form. Thus, for a weak gravitational field [in which case (4.142) is a valid approximation], we expect m/ρ to be a small quantity compared with 1 for physically significant values of ρ; that is, $m/\rho \ll 1$. We may then expand (6.69) to first order in m/ρ,

$$\begin{aligned} ds^2 &\cong \left(1 - \frac{m}{\rho}\right)\left(1 - \frac{m}{\rho}\right) c^2\, dt^2 - \left(1 + \frac{2m}{\rho}\right) d\sigma^2 \\ &\cong \left(1 - \frac{2m}{\rho}\right) c^2\, dt^2 - \left(1 + \frac{2m}{\rho}\right) d\sigma^2 \end{aligned} \tag{6.70}$$

If we compare this with (4.142), we see that we must have within our approximation

$$g_{00} \cong \left(1 - \frac{2m}{\rho}\right) \cong \left(1 + \frac{2\varphi}{c^2}\right) = \left(1 - \frac{2\kappa M}{c^2\rho}\right) \tag{6.71}$$

and we thereby obtain the same result as in Sec. 6.1,

$$m = \frac{\kappa M}{c^2} \tag{6.72}$$

Thus both the original Schwarzschild solution and the above isotropic form lead by a correspondence argument to a consistent identification of the constant of integration m.

6.3 The General Relativistic Kepler Problem and the Perihelic Shift of Mercury

The principal results of the preceding sections are the Schwarzschild solution (6.53) and the identification of the constant of integration m in (6.54) and (6.72). In this section we shall use these results to study the motion of a test particle in a Schwarzschild field, which should directly correspond to planetary motion in the gravitational field of the sun.

This problem is the relativistic analogue of the classical Kepler problem of planetary motion in an inverse-square force field.

As a guide in investigating the relativistic problem, let us recall some of the main features of the classical problem. Kepler's first law states that a planet describes a closed elliptical orbit with the sun at a focal point. However (more realistically), the presence of such small influences as other planets moving in the sun's field causes a perturbation in the motion of a given planet, and the resulting orbit is not precisely elliptic. Indeed, one may think of the actual orbit as a slightly bumpy ellipse which may precess in the plane of motion; that is, the perihelion (point of closest approach to the sun) shifts about and does not always occur at the same angular position.

The fact that the idealized classical orbit is a closed ellipse is a result peculiar to the Newtonian inverse-square law; in fact, Newton himself found that, if the force of gravity were proportional to $1/r^{2+\delta}$ instead of $1/r^2$, then a planetary orbit would not be closed and a perihelic shift of order δ would occur. Indeed, this result was taken to indicate that, since planetary orbits are very nearly closed, the Newtonian inverse-square law must be quite accurate, as in fact it is.

Let us now ask what differences might be expected between the predictions of classical celestial mechanics and general relativistic celestial mechanics. Since Kepler's first law is experimentally verified to be correct to high accuracy, we might expect the relativistic theory merely to add a few bumps to the nearly elliptic orbits and contribute somewhat to perihelic motion. Since angles are much more conveniently measured in astronomy than are distances, it is natural to concentrate on perihelic motion. Conveniently enough, there is, in fact, a well-known discrepancy in classical mechanics concerning the perihelic motion of the planet Mercury. Because of Mercury's high velocity and eccentric orbit, the perihelion position can be accurately determined by observation; the difference between the classically predicted perihelic shift (due to perturbation by other planets) and the observed perihelic shift is 43 seconds of arc per century. Even though this is a very small difference, it is about a hundred times the probable observational error and represents a true discrepancy from the very precise predictions of celestial mechanics which has bothered astronomers since the middle of the last century (Leverrier, 1859).

The first attempt to explain this discrepancy consisted in hypothesizing the existence of a new planet, Vulcan, inside the orbit of Mercury, and much theoretical work was done to predict the position of Vulcan, using the known perturbation on Mercury's orbit. However, careful observation failed to discover the hypothetical planet, and the hypothesis was finally abandoned in 1915 when Einstein used general relativity theory

to explain the observed effect. Let us now proceed to investigate the general relativistic Kepler problem and, as an application, study the motion of planetary perihelia.

We must first decide which radial measure to use, either the r of (6.53) or the ρ of (6.69). A perihelic shift involves the angle between *successive minima* of the radial distance; since $\rho(r)$ is a monotonic function of r outside the Schwarzschild singularity [by virtue of (6.62)], the minima of both r and $\rho(r)$ for a planetary orbit occur at the same angular position, so we can use equally well either the original coordinate r or the isotropic coordinate ρ. It will prove more convenient to use r and the line element (6.53) since the resultant equations of motion will be more similar to the equations of motion for the classical Kepler problem and easier to interpret.

As we stated in Sec. 4.3, the motion of a body in a gravitational field follows a four-dimensional geodesic line. Hence, to find the orbit of a planet, we need the Euler-Lagrange equations for the following variational problem:

$$\delta \int ds = 0 \tag{6.73}$$

where ds is given by the Schwarzschild line element (6.53). As in Sec. 6.1, we may simplify calculations by considering the equivalent variational problem

$$\delta \int \left\{\left(1 - \frac{2m}{r}\right) c^2\dot{t}^2 - \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 - r^2[\dot{\theta}^2 + \sin^2\theta\,\dot{\varphi}^2]\right\} ds = 0 \tag{6.74}$$

(As before, a dot indicates differentiation with respect to s.) The three Euler-Lagrange equations for θ, φ, and t associated with this variational problem are the following:

$$\frac{d}{ds}(r^2\dot{\theta}) = r^2 \sin\theta \cos\theta\,\dot{\varphi}^2 \tag{6.75}$$

$$\frac{d}{ds}(r^2 \sin^2\theta\,\dot{\varphi}) = 0 \tag{6.76}$$

$$\frac{d}{ds}\left[\left(1 - \frac{2m}{r}\right)\dot{t}\right] = 0 \tag{6.77}$$

Note that we have not included the Euler-Lagrange equation for r; it is more convenient to divide the line element (6.53) by ds^2 to obtain a fourth differential equation,

$$1 = \left(1 - \frac{2m}{r}\right) c^2 \dot{t}^2 - \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 - r^2(\dot{\theta}^2 + \sin^2\theta\, \dot{\varphi}^2) \tag{6.78}$$

Using the above four differential equations for t, r, θ, and φ as functions of s, it is possible to obtain and solve the equations of a planetary orbit. In classical mechanics the orbit of a body in a central force field lies in a plane. We can show that the same holds true in the present theory. By an appropriate orientation of the axes we can make $\theta = \pi/2$ and $\dot{\theta} = 0$ at some initial s. Then, from (6.75), it follows that, for all s,

$$\theta = \frac{\pi}{2} \tag{6.79}$$

since the initial conditions determine a unique solution of (6.75), and (6.79) is surely such a solution. Substitution of $\theta = \pi/2$ in (6.76) allows us to integrate (6.76) immediately:

$$r^2\dot{\varphi} = h = \text{const} \tag{6.80}$$

Equation (6.77) integrates to

$$\left(1 - \frac{2m}{r}\right)\dot{t} = l = \text{const} \tag{6.81}$$

Substituting the results (6.79), (6.80), and (6.81) into (6.78), we obtain the following differential equation for $r(s)$:

$$1 = \left(1 - \frac{2m}{r}\right)^{-1} c^2 l^2 - \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 - \frac{h^2}{r^2} \tag{6.82}$$

As in the classical Kepler problem, one can simplify matters by considering r as a function of φ instead of s. Denoting differentiation with respect to φ by a prime, we then have

$$r' = \frac{dr}{d\varphi} = \frac{\dot{r}}{\dot{\varphi}} \tag{6.83}$$

From (6.80) and (6.83) we obtain

$$\dot{r} = \dot{\varphi} r' = \frac{h}{r^2} r' \tag{6.84}$$

The differential equation for $r(\varphi)$ is then obtainable from (6.82):

$$\left(1 - \frac{2m}{r}\right) = c^2 l^2 - \frac{h^2}{r^4} r'^2 - \frac{h^2}{r^2}\left(1 - \frac{2m}{r}\right) \tag{6.85}$$

Following once more the example of the classical Kepler problem, we substitute for the dependent variable

$$r = \frac{1}{u} \tag{6.86}$$

which implies

$$r' = -\frac{u'}{u^2} \tag{6.87}$$

Using these relations, we can convert (6.85) to a differential equation for $u(\varphi)$:

$$(1 - 2mu) = c^2 l^2 - h^2 u'^2 - h^2 u^2 (1 - 2mu) \tag{6.88}$$

This reduces to

$$u'^2 = \left(\frac{c^2 l^2 - 1}{h^2}\right) + \frac{2m}{h^2} u - u^2 + 2mu^3 \tag{6.89}$$

which is immediately integrable:

$$\varphi = \varphi_0 + \int_{u_0}^{u} \frac{du}{\left(\dfrac{c^2 l^2 - 1}{h^2} + \dfrac{2m}{h^2} u - u^2 + 2mu^3\right)^{1/2}} \tag{6.90}$$

This is an *exact solution* to the problem; it expresses the angle φ as an integral of $u = 1/r$, and conversely it gives u as the inverse (implicit) function of φ.

Unfortunately, even though (6.90) is a complete solution to our problem, its form is not particularly enlightening; $u(\varphi)$ is given in implicit form, and the approximate classical form of the trajectory (an ellipse) is not at all evident in (6.90). To make the problem more transparent and to establish a closer connection with the classical Kepler problem (which involves a second-order differential equation), we shall convert the first-order equation (6.89) to a second-order equation by differentiation with respect to φ. We obtain

$$2u'u'' = \frac{2m}{h^2}u' - 2uu' + 6mu^2u' \tag{6.91}$$

One possible solution is then obtained by setting the common factor u' equal to zero:

$$u' = 0 \qquad u = \text{const} \qquad r = \text{const} \tag{6.92}$$

Thus circular motion occurs in relativity theory just as in classical theory. [This could also be inferred from the first-order equation (6.89).] The other possible solution, which is much more interesting, will result from canceling the common factor u' from (6.91):

$$u'' + u = \frac{m}{h^2} + 3mu^2 \tag{6.93}$$

This last equation is quite similar in structure to the orbit equation of the classical Kepler problem. Indeed, for the sake of completeness and comparison, let us recall the derivation of Binet's formula for the motion of a particle of mass m in a central field of force with potential function $mf(r)$. We assume, as before, that the motion takes place in a plane $\theta = 0$. Suppressing a common factor of m in the Lagrangian, we find

$$L = \frac{1}{2}\left[\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\varphi}{dt}\right)^2\right] - f(r) \tag{6.94}$$

whence the differential equations of motion

$$\frac{d^2r}{dt^2} = r\left(\frac{d\varphi}{dt}\right)^2 - f'(r) \qquad r^2\frac{d\varphi}{dt} = H = \text{const} \tag{6.95}$$

where $f'(r) = df/dr$. Consider now the trajectory equation for $r = r(\varphi)$, and also introduce the function

$$u(\varphi) = \frac{1}{r(\varphi)} \tag{6.96}$$

We then have

$$\frac{dr}{dt} = \frac{dr}{d\varphi}\frac{d\varphi}{dt} = r'(\varphi)\frac{H}{r^2} = -Hu'(\varphi) \tag{6.97}$$

where $u'(\varphi) = du/d\varphi$. Hence the first differential equation in (6.95) becomes

$$\frac{d^2r}{dt^2} = -H^2u''(\varphi)u^2 = H^2u^3(\varphi) - f'(r) \tag{6.98}$$

Rearranging this, we obtain

$$u'' + u = \frac{1}{H^2}\frac{f'(r)}{u^2} \tag{6.99}$$

which is Binet's general formula describing the $(1/r, \varphi)$ relation for a central force.

For the special case of a Newtonian potential, $f(r) = -\kappa M/r$, Binet's equation (6.99) becomes

$$u'' + u = \frac{\kappa M}{H^2} \tag{6.100}$$

where H is twice the constant areal velocity:

$$H = r^2\frac{d\varphi}{dt} = \text{const} \tag{6.101}$$

The analogous term in the relativistic equation (6.93) is m/h^2, which, by virtue of (6.72) and (6.80), is explicitly given by

$$\frac{m}{h^2} = \frac{\kappa M}{c^2r^4(d\varphi/ds)^2} = \frac{\kappa M}{c^2r^4(d\varphi/dt)^2(dt/ds)^2} \tag{6.102}$$

Furthermore, we know from Sec. 4.3 that, for slowly moving bodies in weak gravitational fields, $(dt/ds)^2$ is approximately $1/c^2$; substituting this in (6.102), we obtain an approximate form for m/h^2:

$$\frac{m}{h^2} \cong \frac{\kappa M}{r^4(d\varphi/dt)^2} = \frac{\kappa M}{H^2} \tag{6.103}$$

Thus we see that the relativistic equation (6.93) differs from the classical equation (6.100) through the addition of the quadratic term $3mu^2$ and has a slightly different constant term m/h^2. One might furthermore expect the term $3mu^2$ to be small relative to the leading constant term; we may easily verify that this is indeed the case by forming the ratio of it and the constant term m/h^2. This ratio is $3u^2h^2$, which, by virtue of (6.80), is $3r^2\dot{\varphi}^2 \cong 3[r(d\varphi/dt)]^2 \cdot 1/c^2$. The quantity $r(d\varphi/dt)$ is the lateral velocity of the planet (the velocity perpendicular to r), so the above ratio may be written as $3v^2_{\text{lateral}}/c^2$, which is always very small and equal to 7.7×10^{-8} in the case of Mercury. The close similarity between the relativistic equation (6.93) and the classical theory (6.100) is now quite clear.

Equation (6.93) may be interpreted, by the above comments, as the Binet equation of motion in classical mechanics for a field of force with the potential

$$f(r) = \frac{-\kappa M}{r} - \frac{\gamma}{r^3} \qquad \gamma = mH^2 = \kappa M h^2 \tag{6.104}$$

Observe, however, that for different values of H, the indicated modification of Newton's law would be different. Thus this analogy may be helpful in a geometric discussion of the trajectory, but has no real physical significance.

For later argument we bring (6.104) into the form

$$f(r) = -\frac{\kappa M}{r}\left(1 + \frac{h^2}{r^2}\right) \tag{6.104a}$$

If v_l denotes the lateral velocity of the planet, we may use the above approximate equation $u^2h^2 = (1/r^2)h^2 \cong (v_l/c)^2$ and write

$$f(r) \cong -\frac{\kappa M}{r}\left[1 + \left(\frac{v_l}{c}\right)^2\right] \tag{6.104b}$$

Since the planetary orbits are very nearly circular, we may also assume $v_l = 2\pi r/T$, where T is the period of revolution of the planet. Finally, by Kepler's third law, we know that r^3/T^2 is the same for all planets, and hence we have $r(v_l/c)^2 = C$ as common value for all planets. Thus

$$f(r) \cong -\frac{\kappa M}{r}\left[1 + \frac{C}{r}\right] \tag{6.104c}$$

a formula in which the angular velocity of the planet has been eliminated. This formulation of (6.104) will be of value later, when we shall discuss contributions to the perihelic shift of nonrelativistic origin.

Let us now investigate the relativistic equation (6.93) with a view to calculating the perihelion shift. We saw above that the term $3mu^2$ represents a small addition to the classical equations, so let us try a perturbation approach. Define

$$A = \frac{m}{h^2} \cong \frac{\kappa M}{H^2} \tag{6.105}$$

and the small dimensionless quantity

$$\epsilon = 3mA \cong \frac{3\kappa^2M^2}{c^2H^2} \tag{6.106}$$

The relativistic orbit equation then takes the form

$$u'' + u = A + \frac{\epsilon u^2}{A} \tag{6.107}$$

To solve this we assume a solution of the form

$$u(\varphi) = u_0(\varphi) + \epsilon v(\varphi) + O(\epsilon^2) \tag{6.108}$$

and attempt to find $u_0(\varphi)$ and $v(\varphi)$.

Substituting this form for u in the differential equation (6.107), we obtain

$$u_0'' + \epsilon v'' + u_0 + \epsilon v = A + \epsilon u_0^2/A + O(\epsilon^2) \tag{6.109}$$

Equating the zeroth-order terms in ϵ, we have

$$u_0'' + u_0 = A \tag{6.110}$$

which is essentially the classical equation (6.100). The solution is easily checked to be

$$u_0 = A + B\cos(\varphi + \delta) \tag{6.111}$$

where B and δ are arbitrary constants. By an appropriate orientation of the axes we may make δ equal to zero, in which case we obtain the familiar equation of an ellipse,

$$u_0 = A + B\cos\varphi \tag{6.112}$$

Similarly, equating the first-order ϵ terms in (6.109), we obtain

$$\begin{aligned} v'' + v = \frac{u_0^2}{A} &= A + 2B\cos\varphi + \frac{B^2}{A}\cos^2\varphi \\ &= \left(A + \frac{B^2}{2A}\right) + 2B\cos\varphi + \frac{B^2}{2A}\cos 2\varphi \end{aligned} \tag{6.113}$$

Note that we need only a nonhomogeneous solution to this equation since the zeroth-order solution already contains a term $B \cos \varphi$, which is the general solution to the homogeneous equation. Despite the cumbersome appearance of (6.113) it is readily solved; since it is *linear* in v, we may write v as the sum $v = v_a + v_b + v_c$, where v_a, v_b, and v_c are solutions of the equations

$$(6.114) \quad v_a'' + v_a = A + \frac{B^2}{2A} \qquad v_b'' + v_b = 2B\cos\varphi \qquad v_c'' + v_c = \frac{B^2}{2A}\cos 2\varphi$$

that is, we superpose the three solutions (6.114) to get (6.113). The nonhomogeneous solutions to (6.114) are easily checked to be

$$(6.115) \qquad v_a = A + \frac{B^2}{2A} \qquad v_b = B\varphi\sin\varphi \qquad v_c = -\frac{B^2}{6A}\cos 2\varphi$$

so a nonhomogeneous solution to (6.113) is

$$(6.116) \quad v = v_a + v_b + v_c = \left(A + \frac{B^2}{2A}\right) + B\varphi\sin\varphi - \frac{B^2}{6A}\cos 2\varphi$$

Combining this with the zeroth-order solution (6.112), we have the entire solution for the orbit to first order in ϵ:

$$\begin{aligned}(6.117) \quad u &= u_0 + \epsilon v \\ &= \left(A + \epsilon A + \frac{\epsilon B^2}{2A}\right) + \left(B\cos\varphi - \frac{\epsilon B^2}{6A}\cos 2\varphi\right) + \epsilon B\varphi\sin\varphi\end{aligned}$$

Using this solution, we can readily calculate the perihelion shift. Since only the last term is nonperiodic, it is clear that whatever irregularities occur in the perihelion position must be due to this term. To clarify further the effect of the nonperiodic term, note that, to first order in ϵ,

$$(6.118) \quad \cos(\varphi - \epsilon\varphi) = \cos\varphi\cos\epsilon\varphi + \sin\varphi\sin\epsilon\varphi = \cos\varphi + \epsilon\varphi\sin\varphi$$

so the solution may be written as

$$(6.119) \quad u = A + B\cos(\varphi - \epsilon\varphi) + \epsilon\left(A + \frac{B^2}{2A} - \frac{B^2}{6A}\cos 2\varphi\right)$$

In this form the effect of the various terms on the orbit is apparent. The basic elliptical orbit is represented by $A + B\cos\varphi$. The effect of the last term is to introduce small *periodic* variations in the radial distance of the planet. Such effects are difficult to detect, and since they are periodic, they cannot influence the perihelic motion. However, the $\epsilon\varphi$ which appears in the cosine argument does indeed introduce a nonperiodicity, and since φ can become large, the effect is not negligible. Accordingly, let us write (6.119) in the form

$$(6.120) \quad u = A + B\cos(\varphi - \epsilon\varphi) + (\text{periodic terms of order } \epsilon)$$

The perihelion of a planet occurs when r is a minimum or when $u = 1/r$ is a maximum. From (6.120) we see that u is maximum when

$$(6.121) \qquad \varphi(1 - \epsilon) = 2\pi n$$

or approximately

$$(6.122) \qquad \varphi = 2\pi n(1 + \epsilon)$$

Therefore successive perihelia will occur at intervals of

$$(6.123) \qquad \Delta\varphi = 2\pi(1 + \epsilon)$$

instead of 2π as in periodic motion. Thus the perihelion *shift* per revolution is given by

$$(6.124) \qquad \delta\varphi = 2\pi\epsilon = 2\pi\left(\frac{3\kappa^2 M^2}{c^2 H^2}\right)$$

For the case of Mercury, Eq. (6.124) gives a total shift of 43.03″ per century. This is in excellent agreement with the observational result of 43.11 ± 0.45″ which is unaccounted for classically. (For a more extensive discussion of the observational problem, see the review article of Finlay-Freundlich, 1955, and the article of Shapiro, 1972.) This fact is of crucial importance; historically it was the first major observational test of general relativity theory. Recently, however, Dicke (Dicke, 1964) has questioned the excellence of this test by calling into question the exact shape of the sun and hence its classical gravitational field. We discuss this further in the following section.

6.4 The Sun's Quadrupole Moment and Perihelic Motion

In Sec. 6.3 we derived a formula for the perihelic motion of a planet due to the relativistic correction to Newton's law of gravitation and showed that the result is in good agreement with observation in the case of Mercury. We used this fact to strengthen our confidence in Einstein's field equations. Thus it is very important to discuss this effect critically and evaluate all classical contributions.

We mentioned in Sec. 6.3 that attempts were made in the nineteenth century to explain the advance of the perihelion of Mercury by hypothesizing a planet Vulcan between Mercury and the sun, whose influence would cause the discrepancy between the predictions of classical celestial mechanics and the actual motion. This explanation was given up since the planet Vulcan was never observed. However, a very similar explanation might be given for at least part of the effect, namely, that a very small flattening of the solar sphere into an ellipsoid would also lead to perihelic shifts. The perturbation on planetary orbits by a surplus mass located in a ring around the equator of the sun would cause perturbations similar to those of a small planet just grazing the surface of the sun. Such a solar flattening has been measured by Dicke and Goldenberg (Dicke and Goldenberg, 1967), who find a flattening of 5.0 ± 0.4 parts in 10^5. It is somewhat difficult to reconcile this value with the rotation rate of the surface of the sun; it is necessary to make the awkward assumption that the interior of the sun rotates faster than the surface. This leads to some doubt whether the mass distribution of the sun is flattened by the same amount as the visual sphere. If one assumes that it is, however, this measurement indicates that the resultant quadrupole moment contributes about 3.4″ per century to the perihelic shift of Mercury, so that relativity theory and observation would differ by about 8 per cent for this crucial test. However, the measurements of H. Hill (1974) yield a flattening of only 1.0 ± 0.7 parts in 10^5, a result fully consistent with a uniformly rotating sun. It is clearly of interest to study this question further and to evaluate the quadrupole-induced perihelic shift and the relativistic perihelic shift for all the planets.

We shall first calculate the potential created by a sphere which is widened by a bulge around its equator. This potential will depend only on the distance r from the center of the sphere and the azimuthal angle θ from the polar axis. It has to be unchanged if we go from any point in space to its mirror image in the equatorial plane. Hence, if we develop the potential $f(r,\theta)$ in spherical harmonics, the leading two terms will be of the form

$$f(r,\theta) = -\frac{\kappa M}{r}\left[1 + D\,\frac{3\cos^2\theta - 1}{r^2}\right] + O\left(\frac{1}{r^4}\right) \tag{6.125}$$

That is, the usual potential of a sphere has been corrected by the addition of a quadrupole term. The factor D can easily be calculated on the

basis of the details of the deformation, but its precise value is not important for our discussion. Since we assume that the deformation from the spherical form is small and will use the potential relatively far from the center of the sun, we shall neglect the terms $O(1/r^4)$. Since all planets near the sun lie essentially in the plane of the ecliptic where $\theta = \pi/2$, the potential $f(r,\theta)$ acts on them as if it has the purely radial dependence

$$f(r) = -\frac{\kappa M}{r} - \frac{B}{r^3} \tag{6.126}$$

The motion of the planets lies in the equatorial plane and may be described by the relation $r = r(\varphi)$. We have, by Binet's formula (6.99), the following differential equation for the function $u(\varphi) = r(\varphi)^{-1}$ in the field of force (6.126):

$$u'' + u = \frac{1}{u^2H^2}f'(r) = \frac{1}{H^2}(\kappa M + 3Bu^2) = A + \frac{\epsilon}{A}u^2 \tag{6.127}$$

where $H = r^2(d\varphi/dt)$, and we have defined parameters

$$A = \frac{\kappa M}{H^2} \qquad \epsilon = \frac{3\kappa M}{H^4}B \tag{6.128}$$

The dimensionless quantity ϵ may be considered as small since the factor B depends on the deformation of the sphere and is very small. We can therefore use the result of the perturbation theory applied to the formally identical problem (6.107). We found in (6.124) the corresponding perihelic shift per revolution $\delta\varphi = 2\pi\epsilon$, which becomes in the present problem, by virtue of (6.128),

$$\delta\varphi = \frac{6\pi\kappa M}{H^4}B \tag{6.129}$$

Astronomers observe the shift in the perihelic motion after many revolutions of the planet. They express, therefore, the perihelic shift per century by the formula

$$S = \frac{\delta\varphi}{T} = \frac{6\pi\kappa M}{H^4T}B \tag{6.130}$$

where T is the period of revolution expressed in units of centuries. We may also relate the constant H of areal velocity to the period T. If r is the mean distance of the planet from the sun, we may integrate the relation

$$r^2 \frac{d\varphi}{dt} = H \tag{6.131}$$

over one revolution and obtain the approximate relation

$$2\pi r^2 = HT \tag{6.132}$$

(Note that this is not valid for some of the minor planets with a large eccentricity, e.g., Icarus.) Thus (6.130) becomes

$$S = 6\pi\kappa MB \frac{T^3}{(2\pi)^4 r^8} \tag{6.133}$$

Finally, we make use of Kepler's third law of planetary motion, which asserts that

$$\frac{T^2}{r^3} = C \tag{6.134}$$

has the same value for all planets. Hence (6.133) reduces to

$$S = \left(\frac{3\kappa MBC^{3/2}}{(2\pi)^3}\right) r^{-7/2} \tag{6.135}$$

Since only r varies from planet to planet, we recognize that the quadrupole-induced perihelic shift for different planets would vary as the $-\frac{7}{2}$ power of their distance from the sun.

Let us next compare the result with the prediction of the general relativity theory. We have the following formula from Sec. 6.3:

$$\delta\varphi = 2\pi \left(\frac{3\kappa^2 M^2}{c^2 H^2}\right) \tag{6.136}$$

from which follows the perihelic shift per century,

$$S = \frac{6\pi\kappa^2 M^2}{c^2} \frac{1}{H^2 T} \tag{6.137}$$

in which the period T of the planet is expressed in centuries. We derive, by use of (6.132) and (6.134),

$$S = \frac{3\kappa^2 M^2 C^{1/2}}{2\pi c^2} r^{-5/2} \tag{6.138}$$

The perihelic shift as predicted by Einstein's formula thus varies as the $-\frac{5}{2}$ power of the distance of the planet from the sun.

It is very important to note that the r dependence of the relativistic effect is different from that of the quadrupole effect. Thus, in principle, it is possible to evaluate both contributions. For example, since both effects are small perturbations to the usual classical orbit, the total effect may be written as a linear sum

$$S = \lambda \frac{3\kappa^2 M^2 C^{1/2}}{2\pi c^2} r^{-5/2} + \frac{3\kappa MBC^{3/2}}{(2\pi)^3} r^{-7/2} \tag{6.139}$$

where according to relativity theory $\lambda = 1$. Thus it is necessary to measure the shift of two or more planets to obtain an observational value for λ and B; one would thus test relativity theory as well as measure the solar quadrupole moment.

The values of the perihelic shifts as presently known and as calculated from relativity theory are given in Table 6.1. The quadrupole-moment correction suggested by Dicke is not included. The separation of relativistic and quadrupole effects is not feasible at present, but much more precise measurements should be possible with the use of planetary radar reflection (Shapiro et al., 1971). Such measurements should soon provide much more accurate values of the shift for the inner planets.

TABLE 6.1

	Distance r from sun, $\times 10^9$ m	Shift S, seconds of arc/century	
		Calculated	Observed
Mercury†	58	43.03	43.11 ± 0.45
Venus†	108	8.6	8.4 ± 4.8
Earth†	149	3.8	5.0 ± 1.2
Icarus‡	161	10.3	9.8 ± 0.8

† Data from Duncombe (1956).
‡ Data from Shapiro et al. (1971).

6.5 The Trajectory of a Light Ray in a Schwarzschild Field

In this section we shall treat a second interesting case of motion in the sun's gravitational field, the trajectory of a light ray. This problem is particularly interesting because, as with Mercury's perihelion shift, the predictions can be subjected to observational test within the solar system.

In order to treat this problem we need to make two assumptions about the propagation of light rays in a Riemann space: (1) As with the case of a massive test particle, we assume that the trajectory is a geodesic line in a four-dimensional space. (2) In special relativity the path of a light ray (which lieson the light cone) is characterized in space-time by its null line element, $ds^2 = 0$. We assume that the same is true in general relativity. Thus, in short, the light-ray trajectories are null-geodesic lines.

When discussing null geodesics we must observe that the curve parameter s which we have been using until now is no longer admissible since $s=0$ holds on null geodesics. We have to return to the original concept of parallel displacement, i.e., to ask that a null vector dx^α/dq be parallel-displaced in terms of the arbitrary parameter q according to the general law

$$\frac{d}{dq}\left(\frac{dx^\alpha}{dq}\right) + \begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix} \frac{dx^\beta}{dq}\frac{dx^\gamma}{dq} = 0 \tag{6.140}$$

By the general theory this vector will preserve its length; that is, it will remain a null vector. It is easy to see that the above differential equations for the null geodesic are equivalent to the variational problem

$$\delta \int g_{\alpha\beta} \frac{dx^\alpha}{dq}\frac{dx^\beta}{dq}\, dq = 0 \tag{6.141}$$

The parameter q belongs to the family of distinguished parameters discussed in Sec. 2.3. Recall that all parameters of this family are linearly related. In the case of the Schwarzschild metric, we find the equations of motion for φ and t as before [(6.80) and (6.81)]:

$$r^2\dot{\varphi} = \tilde{h} = \text{const} \qquad \left(1 - \frac{2m}{r}\right)\dot{t} = \tilde{l} = \text{const} \tag{6.142}$$

The dots now denote differentiation with respect to q, and we have assumed as before that $\theta = \pi/2$. Instead of Eq. (6.82), however, we now obtain, since $ds^2 = 0$,

$$0 = \left(1 - \frac{2m}{r}\right)^{-1} c^2\tilde{l}^2 - \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 - \frac{\tilde{h}^2}{r^2} \tag{6.143}$$

Thus, proceeding as before with the substitutions $u(\varphi) = 1/r(\varphi)$, we obtain

$$0 = c^2\tilde{l}^2 - \tilde{h}^2u'^2 - \tilde{h}^2u^2(1 - 2mu) \tag{6.144}$$

and by differentiation with respect to φ,

$$u'(u'' + u - 3mu^2) = 0 \tag{6.145}$$

Temporarily discarding the special solution $u = \text{const}$, we finally arrive at the equation for a light-ray trajectory

$$u'' + u = 3mu^2 \tag{6.146}$$

Before continuing with (6.146), let us return for a moment to the formal solution $u = \text{const}$, which entered the theory through differentiation of (6.144). This solution would describe light rays circling the attracting center at a fixed distance $r = r_0$. Such singular solutions occurred also in the theory of planetary motion, and in that theory they have physical reality. The situation in the present case is different. Observe that the general equation (6.93) admits $u = u_0$ as a solution for an appropriate choice of the initial angular momentum. However, $u = u_0$ is a solution of the light-ray equation (6.146) only if $u_0^{-1} = r_0 = 3m$. Hence the singular solutions of $u' = 0$ cannot be changed continuously into solutions of the more general equation (6.146) except at $r_0 = 3m$. Thus these solutions are in general unstable. Indeed, the general equation for light rays should be of second order so that rays through every point and in every direction are possible. This condition is fulfilled by (6.146), but not by $u' = 0$.

It is interesting to note that (6.146) can also be deduced from (6.93) by intuitive reasoning. Equation (6.93) describes the orbit or trajectory of a particle in the Schwarzschild field:

$$u'' + u = \frac{m}{h^2} + 3mu^2 \tag{6.147}$$

Using the expression for m given by (6.72) and the (exact) expression for m/h^2 given by (6.102), we can write this as

$$u'' + u = \frac{\kappa M}{c^2r^4}\left(\frac{ds}{d\varphi}\right)^2 + 3\,\frac{\kappa M u^2}{c^2} \tag{6.148}$$

This equation for the geodesics follows directly from the variational problem (6.74) and involves no approximation. In order to specialize this to the case of a light ray, we must additionally set $ds^2 = 0$. Since the angular interval $d\varphi$ will in general be nonzero as the light ray sweeps by the sun, we conclude that, for the limiting case of a null geodesic,

$$\frac{m}{h^2} = \frac{\kappa M}{c^2 r^4}\left(\frac{ds}{d\varphi}\right)^2 = 0 \tag{6.149}$$

It follows that the equation of the trajectory is, in agreement with our preceding derivation,

$$u'' + u = 3mu^2 \qquad \text{(null geodesic)} \tag{6.150}$$

As with the orbit equation of Sec. 6.3, we can show that the term $3mu^2$ is small relative to the other terms of the equation. To do this, form the ratio of $3mu^2$ to the term u; that is, consider $3mu$. Using the definition of the Schwarzschild radius $r_s = 2m$ (Sec. 6.1), we may also write this ratio as $\frac{3}{2}(r_s/r)$. As we mentioned in Sec. 6.1, the Schwarzschild radius of the sun is of the order of a kilometer; thus, for a trajectory outside the sun's surface, the above ratio is evidently very small. This allows us to regard $3mu^2$ as a small perturbation term in Eq. (6.150). Accordingly, let us call

$$3m = \epsilon \tag{6.151}$$

and write the equation of the light-ray trajectory as

$$u'' + u = \epsilon u^2 \tag{6.152}$$

As in Sec. 6.3, we shall use a standard perturbation approach to treat the above equation; we suppose a solution to (6.152) of the form

$$u = u_0 + \epsilon v + O(\epsilon^2) \qquad \epsilon = 3m \tag{6.153}$$

Substituting this in (6.152), we obtain

$$u_0'' + u_0 + \epsilon v'' + \epsilon v = \epsilon u_0^2 + O(\epsilon^2) \tag{6.154}$$

Equating the zeroth-order terms in ϵ, we have

$$u_0'' + u_0 = 0 \tag{6.155}$$

This has the solution (see Fig. 6.1)

$$u_0 = A \sin(\varphi - \varphi_0) \tag{6.156}$$

which, by an appropriate orientation of axes, may be written without the arbitrary constant φ_0:

$$u_0 = A \sin \varphi \tag{6.157}$$

In terms of the first-order radius $r = 1/u_0$, this becomes

$$r \sin \varphi = \frac{1}{A} \tag{6.158}$$

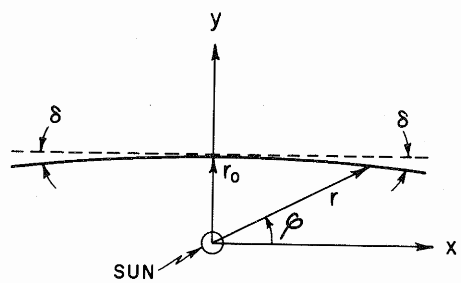
Since $r \sin \varphi$ is simply the Cartesian coordinate y, this evidently represents a straight line parallel to the x axis. This is indeed precisely what we should expect: in first approximation the light ray is not deflected at all by the sun's gravitational field. From Eq. (6.158) it is clear that the distance of closest approach to the origin (the sun) is $1/A$, so we shall call this constant r_0 and write the zeroth-order solution as

$$u_0 = \frac{1}{r_0} \sin \varphi \tag{6.159}$$

Next, equating the first-order ϵ terms of (6.154), we obtain

$$v'' + v = u_0^2 = \frac{1}{r_0^2}\sin^2\varphi = \frac{1}{2r_0^2}(1 - \cos 2\varphi) \tag{6.160}$$

Fig. 6.1
Deflection of light by the sun. The dotted line is the undeflected path $r \sin \varphi = r_0$, and the solid line is the deflected path. δ is the angle between the undeflected path and the asymptote to the deflected path.



To solve this we use a trial solution with unknown coefficients:

$$v = \alpha + \beta \cos 2\varphi \tag{6.161}$$

Differentiation gives

$$v'' = -4\beta \cos 2\varphi \tag{6.162}$$

so that

$$v'' + v = \alpha - 3\beta \cos 2\varphi \tag{6.163}$$

Comparing this term by term with (6.160), we see that (6.161) will be a solution if

$$\alpha = \frac{1}{2r_0^2} \qquad \beta = \frac{1}{6r_0^2} \tag{6.164}$$

Thus a solution of the differential equation (6.160) is

$$v = \frac{1}{2r_0^2} + \frac{1}{6r_0^2} \cos 2\varphi \tag{6.165}$$

Using this and the zeroth-order solution (6.159), we have the full first-order solution to the trajectory equation (6.152):

$$u = \frac{1}{r_0} \sin \varphi + \frac{\epsilon}{2r_0^2} (1 + \tfrac{1}{3} \cos 2\varphi) \tag{6.166}$$

As we have seen above, the trajectory of a light ray as given by (6.166) is essentially a straight line $[u = (1/r_0) \sin \varphi]$ with a perturbation of order ϵ. The effect of this perturbation will alter the trajectory to produce a small overall deflection; that is, light approaches the sun along an asymptotic straight line, is deflected by the gravitational field, and recedes again on another asymptotic straight line. The total deflection can be measured observationally for the case of starlight grazing the sun and arriving finally on the earth. Let us therefore see what total deflection is predicted by (6.166) for such a situation.

The asymptotes of the trajectory will clearly correspond to those values of the angle φ for which r becomes infinite or (equivalently) u becomes zero in (6.166). These asymptotes are nearly parallel to the x axis and correspond to φ being close to zero or π. Thus considering the asymptote near $\varphi = 0$ first and calling δ the small angle between it and the x axis, we approximate $\sin \varphi$ by δ and $\cos 2\varphi$ by 1. Then, setting $u = 0$ in (6.166), we obtain

$$0 = \frac{1}{r_0}\delta + \frac{4}{3}\frac{\epsilon}{2r_0^2} \tag{6.167}$$

or

$$\delta = -\frac{2\epsilon}{3r_0} = -\frac{2m}{r_0} \tag{6.168}$$

The minus sign indicates the light ray is bent inward by the sun. A similar procedure for the other asymptote, for which φ is taken to be $\pi - \delta$, yields the same value, $\delta = -2m/r_0$. Thus the total deflection of the light ray, the angle between the asymptotes, is

$$\Delta = \frac{4m}{r_0} = \frac{4\kappa M}{c^2 r_0} \tag{6.169}$$

For a light ray which just grazes the sun, Eq. (6.169) predicts a deflection of 1.75″. The early attempts to compare this prediction with observational data utilized photographs taken during solar eclipses. The positions of stellar images near the sun during an eclipse were compared with the positions 6 months later, with the sun no longer in the field of view. This procedure is inherently difficult since very small displacements of the images have to be measured. As a result the observational results obtained have ranged from 1.5″ to nearly 3″ (von Klüber, 1960).

With the advent of large radio telescopes and the discovery of the pointlike sources of intense radio emission called *quasars* the deflection can now be measured using long-base-line interferometric techniques when such a source passes near the sun. Measurements range from 1.57 to 1.82″, each with an accuracy of about 0.2″. It should be possible in time to reduce this error to about 0.01″ and obtain an extremely accurate test of the theory (Sramek, 1971).

6.6 Travel Time of Light in a Schwarzschild Field

Another interesting problem concerning the behavior of light in a Schwarzschild field is the question of travel time between two given points. Because space-time is curved in the presence of a gravitational field, this travel time is greater than it would be in flat space, and the difference can be tested experimentally.

We can easily calculate the time delay. It is simple to show that the curvature of the path, as discussed in the preceding section, makes a negligible contribution to the time delay when one considers a ray of light traveling between the earth and another planet. Thus we can approximate the path by a straight line, which we choose to be parallel to the x axis; $r \sin \varphi = r_0$, $\theta = \pi/2$. Clearly r_0 is the distance of closest approach to the sun (see Fig. 6.1). The relationship between the time and space coordinates along the world-line of a light ray is given by setting the Schwarzschild line element to zero:

$$0 = \left(1 - \frac{2m}{r}\right) c^2\, dt^2 - \frac{dr^2}{(1 - 2m/r)} - r^2\, d\varphi^2 \tag{6.170}$$

From the equation of the path of the ray we can reexpress $r^2\, d\varphi^2$ in terms of r and dr, so that we obtain

$$\begin{aligned} c^2\, dt^2 &= \frac{dr^2}{(1 - 2m/r)^2} + \frac{r_0^2\, dr^2}{(1 - 2m/r)(r^2 - r_0^2)} \\ &= \frac{dr^2\,(1 - 2mr_0^2/r^3)}{(1 - r_0^2/r^2)(1 - 2m/r)^2} \end{aligned} \tag{6.171}$$

We now take the square root of this, expand to obtain $c\, dt$ to first order in m,

$$c\, dt = \frac{dr}{\sqrt{1 - r_0^2/r^2}} \left(1 + \frac{2m}{r} - \frac{mr_0^2}{r^3}\right) \tag{6.172a}$$

then integrate

$$\begin{aligned} ct = {} & \left(\sqrt{r_p^2 - r_0^2} + \sqrt{r_e^2 - r_0^2}\right) \\ & + 2m \log \frac{\left(\sqrt{r_p^2 - r_0^2} + r_p\right)\left(\sqrt{r_e^2 - r_0^2} + r_e\right)}{r_0^2} \\ & - m \left(\frac{\sqrt{r_p^2 - r_0^2}}{r_p} + \frac{\sqrt{r_e^2 - r_0^2}}{r_e}\right) \end{aligned} \tag{6.172b}$$

The integration is taken from $r = r_0$ to r_p, the planet radius, and from $r = r_0$ to r_e, the earth radius. It is evident that the first term above is the flat-space result for the earth-planet distance, while the other two terms represent an effective increase in the distance. For the solar system we may regard r as a very reasonable radial coordinate and t as an approximate physical time. Note that, as may be expected, the main contribution to the increase in travel time comes from the part of the trajectory closest to the sun, i.e., for small values of r as is evidenced in the terms proportional to m in the integrand (6.172a).

The experimental verification of the delay has been carried out by sending pulsed radar signals from the earth to Venus and Mercury and timing the echoes as the positions of earth and the planet change relative to the sun. For Venus near superior conjunction the measured delay amounts to about 200 μs. The measurements are within 5 per cent of the calculated delays. These measurements constitute the first entirely new test of general relativity in over 50 years (Shapiro, 1972).

6.7 Null Geodesics and Fermat's Principle

We may obtain an interesting interpretation of our principle that light rays travel along null geodesics in space-time and connect it with a well-known theorem of classical optics. We assume a line element that is time-independent or, in invariant language, stationary. The spatial coordinates will be denoted by x^i and the time coordinate by t. The path of a light ray is then characterized by the following two conditions for a null geodesic:

$$\delta \int \mathfrak{L}(x^i, \dot{x}^i, \dot{t})\, dq = 0 \tag{6.173}$$

and

$$ds^2 = A^2 c^2\, dt^2 + g_{ik}\, dx^i\, dx^k = 0 \tag{6.174a}$$

$$\mathfrak{L}(x^i, \dot{x}^i, \dot{t}) = A^2 c^2 \left(\frac{dt}{dq}\right)^2 + g_{ik} \frac{dx^i}{dq} \frac{dx^k}{dq} = 0 \tag{6.174b}$$

where a dot denotes differentiation with respect to q.

We wish now to compare the integral

$$J = \int \mathfrak{L}(x^i, \dot{x}^i, \dot{t})\, dq \tag{6.175}$$

along the actual light trajectory with the same integral taken over an arbitrary trajectory in space-time which is near the null geodesic, has the same endpoints P_1 and P_2 in the three-space, and satisfies the condition of light velocity $\mathfrak{L}(x^i, \dot{x}^i, \dot{t}) = 0$. There are many nonstationary curves between P_1 and P_2 for which the condition (6.174) is fulfilled. They may start at different moments t_1 at P_1 and end at different moments t_2 at P_2.

On the one hand, it is evident that the integral (6.175) is zero for all competing trajectories since its integrand is identically zero. On the other hand, we can calculate to first order the formal change of the integral near the geodesics; it will not necessarily be formally zero, because (6.175) refers only to competing trajectories which have common endpoints in time *and* space. We find, formally,

$$\delta \int \mathfrak{L}(x^i,\dot{x}^i,\dot{t})\, dq = \int \left[\frac{\partial \mathfrak{L}}{\partial x^i}\,\delta x^i + \frac{\partial \mathfrak{L}}{\partial \dot{x}^i}\,\delta \dot{x}^i + \frac{\partial \mathfrak{L}}{\partial \dot{t}}\,\delta \dot{t} \right] dq \tag{6.176}$$

and by integration by parts

$$\delta \int \mathfrak{L}\, dq = \left[\frac{\partial \mathfrak{L}}{\partial \dot{t}}\,\delta t \right]_{t_1,P_1}^{t_2,P_2} + \int \left\{ \left[\frac{\partial \mathfrak{L}}{\partial x^i} - \frac{d}{dq}\left(\frac{\partial \mathfrak{L}}{\partial \dot{x}^i}\right) \right] \delta x^i - \frac{d}{dq}\left(\frac{\partial \mathfrak{L}}{\partial \dot{t}}\right) \delta t \right\} dq \tag{6.177}$$

since by assumption the δx^i vanish at the endpoints. Observe that the integrand vanishes by the Euler-Lagrange equations which characterize the null geodesic. We refer specially to the last term, for which

$$\frac{d}{dq}\left(\frac{\partial \mathfrak{L}}{\partial \dot{t}}\right) = 0 \qquad \frac{\partial \mathfrak{L}}{\partial \dot{t}} = l = \text{const} \tag{6.178}$$

This is indeed the Euler-Lagrange equation for $t(q)$ since $\partial\mathfrak{L}/\partial t \equiv 0$. Thus (6.177) reduces to

$$\delta \int \mathfrak{L}\, dq = l[\delta t]_{t_1}^{t_2} = l\,\delta T \tag{6.179}$$

where δT is the change in travel time between P_1 and P_2 on the competing path, which allows the correct velocity of light as determined by $\mathfrak{L} = 0$.

Thus, comparing our two ways of computing δJ, we arrive at the result: *The actual path of light in three-space between two given endpoints makes the travel time of light with the prescribed local velocity a stationary value among all admissible paths.* This is the well-known Fermat principle of optics, which we now see is a consequence of our null-geodesic principle.

6.8 The Schwarzschild Radius, Kruskal Coordinates, and the Black Hole

We have noted that in the Schwarzschild line element (6.53) a singularity occurs at $r = 2m$, the Schwarzschild radius; at this radius g_{11} is infinite while g_{00} is zero. Because g_{00} is zero, the spherical surface at

$r = 2m$ is an infinite red shift surface, as is clear from our discussion of the red shift in Secs. 4.2 and 4.4. That is, since light emitted by a radiating atom situated on this surface would be red-shifted to zero frequency as it traveled to larger radii, the atom could not be observed.

When r becomes less than $2m$, the signs of the metric components g_{00} and g_{11} change, g_{11} becoming positive and g_{00} becoming negative. This forces us to reconsider the physical meaning of t and r as time and radial markers inside the Schwarzschild radius. Indeed a world-line along the t axis (r, θ, φ constant) has $ds^2 < 0$ and is a *spacelike* curve, while a world-line along the r axis has $ds^2 > 0$ and is a *timelike* curve. It would thus appear natural to reinterpret r as a time marker and t as a radial marker for events which occur inside the Schwarzschild radius. Since we interpret ds/c to represent the proper time along the world line of a particle, as in Sec. 4.2, we see that ds^2 must be positive along such a path. Thus a massive particle could not remain at a constant value of r inside the Schwarzschild radius since that would imply that $ds^2 < 0$ along its world-line.

These features show that $r = 2m$ is an unusual radius, but it does not follow that the intrinsic space-time geometry becomes singular at $r = 2m$. Indeed the "singularity" is associated with the choice of coordinates. This is indicated by the fact that the *invariants* $R^\mu{}_\mu$ and $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ remain finite at the Schwarzschild radius, as does the determinant of the metric tensor. The Riemann tensor $R^\alpha{}_{\beta\gamma\delta}$ is also finite, in particular the terms $R^i{}_{0j0}$, which correspond to relative Newtonian forces in (5.147) (see Exercises 6.2 to 6.4). We shall investigate the mathematical and physical nature of the Schwarzschild singularity in this section by studying the behavior of a falling test body and by introducing coordinates (Kruskal, 1960) in which no singularity occurs in the metric. Specifically we shall show that the Schwarzschild coordinates r and t are well suited for describing the Schwarzschild geometry in the region of greatest physical interest, $2m < r < \infty$ and $-\infty < t < \infty$, but that an alternative choice can shed light on the nature of the surface and interior of the Schwarzschild sphere at $r = 2m$.

Consider the simple case of inward radial motion in a Schwarzschild geometry. The Euler-Lagrange equations of motion for r and t as functions of the arc length s have been obtained in Sec. 6.3. Equations (6.81) and (6.82) specialized to radial motion, $h = 0$, are

$$\dot{t} = \frac{1}{c}\left(1 - \frac{2m}{r}\right)^{-1} \qquad \dot{r}^2 = \frac{2m}{r} \tag{6.180}$$

We have chosen initial conditions so $l = 1/c$ in (6.81), so that $t = s/c$ at large r, as in special relativity. (See Sec. 7.9 for more on this.) Two descriptions of the radial motion are of interest; one is the radial position as a function of the proper time of the test body s/c, and the other is the radial position as a function of the coordinate time t. Since the proper time is that measured by an observer falling with the body while the coordinate time corresponds to the time measured by an observer at rest a large distance from the central mass, the physical significance of these two descriptions is apparent.

Let us solve first for r as a function of s. From (6.180) we obtain immediately

$$\frac{2}{3\sqrt{2m}}\left(r^{3/2} - r_0^{3/2}\right) = s_0 - s \tag{6.181}$$

erreur, c'est : $(r - r_0)^{3/2}$

where r_0 is the initial position at s_0. Curiously this is the same as the analogous *classical* result! No singular behavior occurs at the Schwarzschild radius, and the body falls continuously to $r = 0$ in a finite proper time, as shown in Fig. 6.2.

To describe the motion in terms of coordinate time t we form dr/dt from (6.180),

$$\frac{dr}{dt} = \frac{\dot{r}}{\dot{t}} = -c\sqrt{\frac{2m}{r}}\left(1 - \frac{2m}{r}\right) \tag{6.182}$$

and integrate to obtain

$$c(t_0 - t) = \frac{2}{3\sqrt{2m}}\left(r^{3/2} - r_0^{3/2} + 6m\sqrt{r} - 6m\sqrt{r_0}\right) - 2m \log \frac{(\sqrt{r} + \sqrt{2m})(\sqrt{r_0} - \sqrt{2m})}{(\sqrt{r_0} + \sqrt{2m})(\sqrt{r} - \sqrt{2m})} \tag{6.183}$$

This result is substantially different from (6.181). For situations where r and r_0 are much larger than $2m$ the two results are approximately the same, as should be expected, while for r very near $2m$ we have asymptotically

$$r - 2m = 8me^{-c(t-t_0)/2m} \tag{6.184}$$

It is thus apparent that $r = 2m$ is approached but never passed by the falling test body if one uses t as a time label. The nature of this radial motion is illustrated in Fig. 6.2 for a situation where r_0 is several times $2m$.

It is clear from the preceding that a description of the Schwarzschild geometry in terms of the coordinates r and t is limited: using proper time, we may study events that, in effect, occur *after* $t = \infty$. Certainly the coordinate t is very useful and physically meaningful since it corresponds to the proper time of an observer at rest far away from the central body. Thus in the finite proper time in which a test body falls to $r = 2m$ we would expect that the entire evolution of the physical universe exterior to $r = 2m$ has already occurred, so that the physical meaning of further fall becomes questionable in the context of Schwarzschild geometry. Nevertheless, we can still pose the purely mathematical question of how the fall proceeds in terms of coordinates that have a wider range of usefulness than the Schwarzschild coordinates. We shall therefore study the Kruskal coordinates in order to describe the entire Schwarzschild geometry conveniently.

To introduce Kruskal's coordinates we observe a peculiar feature of light propagation in terms of Schwarzschild coordinates. World-lines of light rays are characterized by a vanishing line element, $ds^2 = 0$. Thus for radial motion the path of a light ray is such that

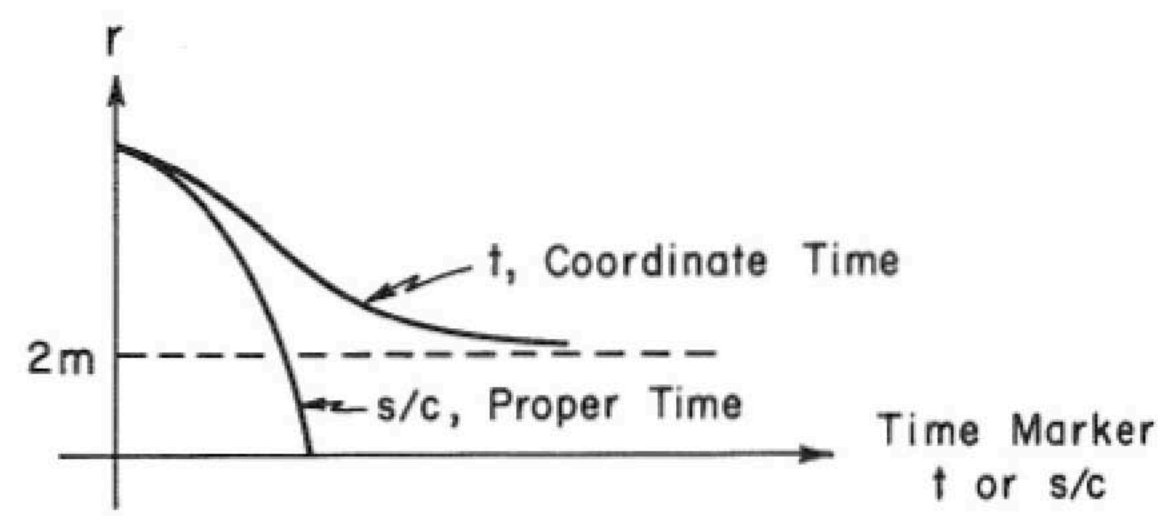
$$\left(1 - \frac{2m}{r}\right)c^2\,dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 = 0 \tag{6.185}$$

This means that the coordinate velocity of light is given by

$$\left(\frac{dr}{dt}\right)^2 = c^2\left(1 - \frac{2m}{r}\right)^2 \tag{6.186}$$

so that at $r = 2m$ the radial coordinate velocity of light becomes zero. This is an undesirable feature of the Schwarzschild coordinates that we can eliminate as follows; we seek a transformation from r and t to new

Fig. 6.2
Fall toward the origin of a Schwarzschild geometry in terms of coordinate time t and proper time on the test body s/c.



variables u and v in which the line element has the form

$$ds^2 = f^2(u,v)(dv^2 - du^2) - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2) \tag{6.187}$$

By the same procedure as above we find the radial coordinate velocity of light to be unity everywhere

$$\left(\frac{du}{dv}\right)^2 = 1 \tag{6.188}$$

so long as f^2 has no zeros. Thus in the u, v coordinates no natural boundary to light propagation can occur.

It is a simple task to obtain from (6.187) differential equations which lead to a transformation from r, t to u, v coordinates and a nonzero function f. The angular coordinates θ and φ will not be changed. The fundamental transformation equation for the metric tensor,

$$g_{\alpha\beta} = \frac{\partial\hat{x}^\mu}{\partial x^\alpha}\frac{\partial\hat{x}^\nu}{\partial x^\beta}\hat{g}_{\mu\nu} \tag{6.189}$$

and the line elements (6.53) and (6.187) lead to the following differential equations to be solved

$$\begin{gathered}1 - \frac{2m}{r} = f^2\left[\left(\frac{\partial v}{\partial x^0}\right)^2 - \left(\frac{\partial u}{\partial x^0}\right)^2\right] \\ -\left(1 - \frac{2m}{r}\right)^{-1} = f^2\left[\left(\frac{\partial v}{\partial r}\right)^2 - \left(\frac{\partial u}{\partial r}\right)^2\right] \qquad x^0 = ct \\ 0 = \frac{\partial u}{\partial x^0}\frac{\partial u}{\partial r} - \frac{\partial v}{\partial x^0}\frac{\partial v}{\partial r}\end{gathered} \tag{6.190}$$

Note that the signs of u and v are not determined by these equations. To simplify we introduce a new radial parameter ξ and a function $F(\xi)$ by

$$\begin{gathered}\xi = r + 2m\log\left|\frac{r}{2m} - 1\right| \\ F(\xi) = \frac{1 - 2m/r}{f^2(r)}\end{gathered} \tag{6.191}$$

We have here assumed that a function f may be found which depends only on r; this is a critical point since an infinite number of transformations could lead to the metric form (6.187), and only this assumption leads to the Kruskal form and also removes the coordinate singularity

at $r = 2m$. The relations (6.190) now simplify to

$$\left(\frac{\partial v}{\partial x^0}\right)^2 - \left(\frac{\partial u}{\partial x^0}\right)^2 = F(\xi) \tag{6.192a}$$

$$\left(\frac{\partial v}{\partial \xi}\right)^2 - \left(\frac{\partial u}{\partial \xi}\right)^2 = -F(\xi) \tag{6.192b}$$

$$\frac{\partial u}{\partial x^0}\frac{\partial u}{\partial \xi} = \frac{\partial v}{\partial x^0}\frac{\partial v}{\partial \xi} \tag{6.192c}$$

If we add Eqs. (6.192a) and (6.192b) and alternately add or subtract twice (6.192c), we obtain

$$\left(\frac{\partial v}{\partial x^0} + \frac{\partial v}{\partial \xi}\right)^2 = \left(\frac{\partial u}{\partial x^0} + \frac{\partial u}{\partial \xi}\right)^2 \tag{6.193a}$$

$$\left(\frac{\partial v}{\partial x^0} - \frac{\partial v}{\partial \xi}\right)^2 = \left(\frac{\partial u}{\partial x^0} - \frac{\partial u}{\partial \xi}\right)^2 \tag{6.193b}$$

Using a relative plus sign for the roots of (6.193a) and a relative minus sign for the roots of (6.193b), we then obtain two equations (if we were to use the same sign, the Jacobian of the transformation would vanish):

$$\frac{\partial v}{\partial x^0} = \frac{\partial u}{\partial \xi} \qquad \frac{\partial v}{\partial \xi} = \frac{\partial u}{\partial x^0} \tag{6.194}$$

which lead to

$$\frac{\partial^2 u}{\partial x^{02}} - \frac{\partial^2 u}{\partial \xi^2} = 0 \qquad \frac{\partial^2 v}{\partial x^{02}} - \frac{\partial^2 v}{\partial \xi^2} = 0 \tag{6.195}$$

Thus both u and v satisfy the wave equation in x^0 and ξ. [If we had chosen the opposite roots in (6.193), the same equation (6.195) would have resulted.]

The general solution of the wave equation is an arbitrary twice-differentiable function of $x^0 \pm \xi$. Thus the solutions to (6.194) and (6.195) are easily seen to be

$$\begin{aligned}v &= h(\xi + x^0) + g(\xi - x^0) \\ u &= h(\xi + x^0) - g(\xi - x^0)\end{aligned} \tag{6.196}$$

where h and g are to be determined. Now we substitute u and v from (6.196) back into Eqs. (6.192); Eq. (6.192c) is automatically satisfied, while (6.192*a*) and (6.192*b*) are equivalent and lead to

$$-4h'(\xi + x^0)g'(\xi - x^0) = F(\xi) \tag{6.197}$$

where a prime indicates differentiation with respect to the argument. This is a remarkable equation that will lead to solutions for h, g, and F that are unique up to unimportant constants.

So far we have made no restrictions on the range of r in our transformation. Now we must specify whether r is greater than or less than $2m$, since in the two regions we shall have different transformation functions which must be patched together at the boundary. We first consider $r \geq 2m$, in which case F is positive from (6.191). To solve (6.197) we differentiate with respect to ξ and x^0 to obtain

$$\frac{F'(\xi)}{F(\xi)} = \frac{h''(\xi + x^0)}{h'(\xi + x^0)} + \frac{g''(\xi - x^0)}{g'(\xi - x^0)} \tag{6.198a}$$

$$0 = \frac{h''(\xi + x^0)}{h'(\xi + x^0)} - \frac{g''(\xi - x^0)}{g'(\xi - x^0)} \tag{6.198b}$$

Thus

$$[\log F(\xi)]' = 2[\log h'(\xi + x^0)]' \tag{6.199}$$

We may treat ξ and $y \equiv \xi + x^0$ as independent variables, which implies that the two sides of (6.199) are functions of two independent variables and must both be equal to some constant η. Thus from (6.199) and (6.198*b*) we see that h, g, and F are exponential functions. We therefore write the solution to (6.197) as

$$h(y) = \tfrac{1}{2}e^{\eta y} \qquad g(y) = -\tfrac{1}{2}e^{\eta y} \qquad F(\xi) = \eta^2 e^{2\eta\xi} \tag{6.200}$$

where the arbitrary additive constants are chosen to be zero and the multiplicative constants to be $\frac{1}{2}$ for convenience. Note that the relative sign of h and g is negative, as dictated by $F > 0$. Now from (6.191), (6.196), and (6.200) we have the transformation

$$\begin{aligned} u &= \left(\frac{r}{2m} - 1\right)^{2m\eta} e^{\eta r} \cosh \eta x^0 \\ v &= \left(\frac{r}{2m} - 1\right)^{2m\eta} e^{\eta r} \sinh \eta x^0 \\ f^2 &= \frac{2m}{\eta^2 r}\left(\frac{r}{2m} - 1\right)^{1-4m\eta} e^{-2\eta r} \end{aligned} \tag{6.201}$$

It remains only to choose the arbitrary parameter η; to do this we demand that f^2 have no zero or singularity at $r = 2m$, which requires that $\eta = 1/4m$. (Then ds^2 will vanish *only* on the light cone.) The transformation is thus, finally,

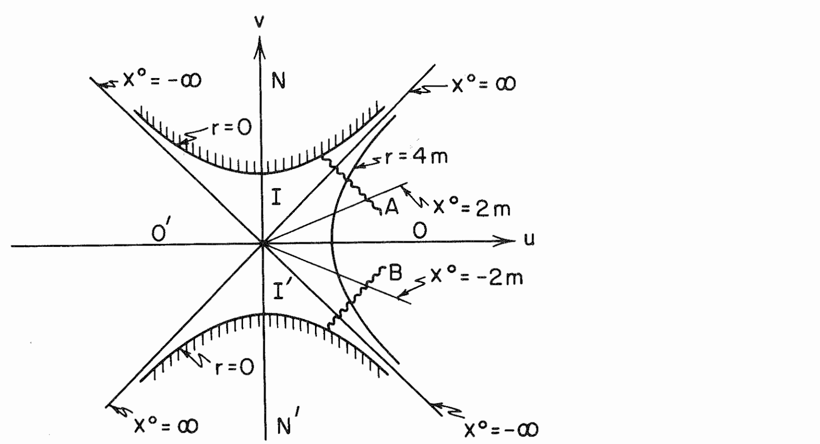
$$\begin{aligned} u &= \sqrt{\frac{r}{2m} - 1}\, e^{r/4m} \cosh \frac{x^0}{4m} \\ v &= \sqrt{\frac{r}{2m} - 1}\, e^{r/4m} \sinh \frac{x^0}{4m} \\ f^2 &= \frac{32m^3}{r} e^{-r/2m} \end{aligned} \tag{6.202}$$

The region of the uv plane defined by (6.202) for $r \geq 2m$ is $u \geq |v|$, which is labeled O in Fig. 6.3. Some special lines are of interest; for any finite x^0 the boundary line $r = 2m$ in the rt plane corresponds to the point $u = v = 0$ in the uv plane. Also we note that $x^0 \to \infty$ corresponds to $u = v$ and $x^0 \to -\infty$ corresponds to $u = -v$ for any value of $r > 2m$. For other points in O we invert the transformation (6.202)

$$u^2 - v^2 = \left(\frac{r}{2m} - 1\right) e^{r/2m} \qquad \frac{v}{u} = \tanh \frac{x^0}{4m} \tag{6.203}$$

Thus lines of constant r and lines of constant x^0 form a mesh of intersecting hyperbolas and rays in O as shown in Fig. 6.3. As r approaches

Fig. 6.3
Kruskal coordinates with several lines of constant r and t shown. The regions O and O' correspond to $r > 2m$, while I and I' correspond to $r < 2m$.



$2m$, the hyperbolas approach the lines $u = |v|$, but for $r = 2m$ the hyperbola degenerates to the point $u = v = 0$.

In Eq. (6.200) we arbitrarily chose h to be positive, so that g was negative. We could just as well have chosen the opposite signs, which would then have reversed the signs of both u and v in (6.202). We can therefore *identify* $-u$, $-v$ with u, v so that the nonoverlapping regions O and O' in the uv plane both represent the exterior of the Schwarzschild radius; we must still show that this is consistent with the transformation in the remainder of the uv plane.

For $r < 2m$ we must repeat some of our derivation. From (6.191) we see that F is negative, so that the relative sign of g and h must be positive. Proceeding as before with g and h both positive, we arrive at a transformation appropriate to the interior of the Schwarzschild radius

$$\begin{aligned} u &= \sqrt{1 - \frac{r}{2m}}\, e^{r/4m} \sinh \frac{x^0}{4m} \\ v &= \sqrt{1 - \frac{r}{2m}}\, e^{r/4m} \cosh \frac{x^0}{4m} \\ f^2 &= \frac{32m^3}{r}\, e^{-r/2m} \end{aligned} \tag{6.204}$$

This transformation relates $r < 2m$ to the region $v > |u|$, labeled I in Fig. 6.3. It is important to note that the transformations (6.204) and (6.202) match at the boundary $x^0 = \infty$ and, with appropriate sign changes, at $x^0 = -\infty$. The transformations (6.202) and (6.204) are therefore consistent. The inverse of (6.204) is

$$\begin{aligned} v^2 - u^2 &= \left(1 - \frac{r}{2m}\right) e^{r/2m} \\ \frac{u}{v} &= \tanh \frac{x^0}{4m} \end{aligned} \tag{6.205}$$

so we have another mesh of hyperbolas and rays in I representing constant values of r and t. Note that the hyperbolas (6.203) and (6.205) are the *same*, and so our results for O and I are consistent as we approach the boundary $u = v$ from both sides. The origin $r = 0$ maps onto the hyperbola $v^2 - u^2 = 1$. As discussed for region O' we can identify points u, v in I' with $-u$, $-v$ in I. This is again consistent with values on the boundary curves.

In Kruskal's coordinates the geometry represented by the line element (6.187) with f^2 given in (6.202) is a solution of the Einstein equations and is nonsingular almost everywhere. Only along the hyperbola $v^2 - u^2 = 1$, corresponding to $r = 0$, do singularities occur in the Riemann tensor. Moreover, as we first demanded, light rays always travel along straight lines, $(du/dv)^2 = 1$, as in special relativity. Thus u serves as a global radial marker, and v serves as a global time marker. They do not, however, correspond to spherical coordinates for flat space at asymptotic distances, as the Schwarzschild coordinates do. We must also demand that Eq. (6.203) or (6.205) define r uniquely as an implicit function of u and v. This will be so as long as the right-hand side of (6.205) is a monotonic function of r. Since its derivative is $-(r/2m)e^{r/2m}$, it is monotonic for $r > 0$. Only at $r = 0$ does the derivative vanish and the above result break down; since this corresponds to $v^2 - u^2 = 1$, we see that the shaded regions N and N' in Fig. 6.3 must be deleted from the admissible regions of the Kruskal diagram.

The Schwarzschild metric is a nonsingular solution of the Einstein equation in O, O', I, and I' but has a singularity at the boundary corresponding to $r = 2m$. The Kruskal metric is also a nonsingular solution of the Einstein equations in these regions and is equivalent to the Schwarzschild solution, but it has no singularity at the boundary. This result is analogous to analytic continuation in complex-function theory and is appropriately referred to as an *analytic extension* of the manifold.

Let us use the Kruskal coordinates to study a light ray traveling radially inward toward the Schwarzschild radius, a problem we have already treated in Schwarzschild coordinates. Such a ray is represented by the line A in Fig. 6.3; it has slope -1. In terms of u, v the trajectory is simple; in terms of r and t, however, we see that it begins at some finite $r > 2m$ and finite x^0, travels inward toward $r = 2m$ as $x^0 \to \infty$, and *crosses* the line $x^0 = \infty$ to the interior of the Schwarzschild sphere. After that r continues to decrease along the trajectory, but x^0 *decreases*. This is in agreement with the behavior we previously obtained using Schwarzschild coordinates but goes beyond it and describes the trajectory subsequent to crossing the $x^0 = \infty$ line. The present treatment also clarifies the fact that x^0 is not a reasonable time marker inside the Schwarzschild sphere. For rays emitted from the interior of the Schwarzschild sphere the Kruskal coordinates remain useful. For example, consider a ray B, emitted from $r = 0$. It travels through increasing values of r but decreasing values of x^0, then crosses the line $x^0 = -\infty$ to the exterior, where its evolution is normal.

From the above it is clear that incoming light will in effect be totally absorbed by the Schwarzschild sphere. Since light emerging from the sphere must have been traveling since $x^0 = -\infty$, in effect before the beginning of time, it is questionable whether such light could be observed. If not, then the surface would have the physical aspect of a black hole, a surface that absorbs all light and emits none. It is evident also, a fortiori, that the same is true of massive bodies, since they move within the light cone, i.e., have a slope $(dv/du)^2 > 1$. In Chap. 7 we shall give a more general discussion of the nature and operation of this "one-way membrane."

We have discussed an ideal Schwarzschild black hole. During most of its lifetime the size of a star is determined by a balance between the inward pull of gravity and the pressure due to radiation released by the nuclear-fusion reaction in the deep interior. When the necessary light nuclei have been used up, the fusion process ceases, the stellar equilibrium cannot be maintained, and in some cases the gravitational force collapses the star, shrinking it to a size asymptotically approaching its Schwarzschild radius. Thus for radii slightly greater than $2m$ and times that are finite the geometry becomes that which we have discussed: the asymptotically collapsing star would appear as a black hole. Note that in this case the line $x^0 = -\infty$ is not part of the accessible exterior and photons such as B in Fig. 6.3 would certainly not be expected to be emitted by a real star approaching the black-hole state asymptotically.

Clearly our considerations apply only to nonrotating stars. In the next chapter we discuss black holes with rotation, and in Chap. 14 we discuss terminal stellar evolution and the formation of black holes.

Exercises

6.1 Verify that the off-diagonal elements of the contracted Riemann tensor for a Schwarzschild *form* of line element (6.9) are identically zero, as noted following Eq. (6.52).

6.2 What is $\sqrt{-g}$ for the Schwarzschild metric? Where is it singular?

6.3 Calculate the Riemann tensor for the Schwarzschild metric. Where does it have singularities? What are the components $R^i{}_{0k0}$, which correspond to derivatives of the Newtonian force in (5.147), and where are they singular? (These force derivatives correspond to *tidal forces*.)

6.4 Calculate the invariant $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ for the Schwarzschild metric. Where is it singular? Calculate the scalar curvature for the three-space part of the Schwarzschild metric, $R^i{}_i$. Where is it singular?

6.5 Verify that $\xi_\alpha = (1,0,0,0)$ is a Killing vector in the Schwarzschild geometry. What are the Killing vectors that correspond to rotational symmetry?

6.6 Show that the Petrov type of the Schwarzschild metric is *ID*. (This calculation is much easier than might be expected; why?)

6.7 Show that the effect of bending of the trajectory on the time delay for the radar-echo experiment is negligible.

6.8 Suppose that light were composed of corpuscules that behave like ordinary Newtonian particles in a classical gravitational field. They would then be distinguished only by their asymptotic velocity c in the region far from gravitational influence. Show that with this model one obtains one-half the Einstein value for the deflection of starlight by the sun.

6.9 With the Newtonian corpuscular model of Exercise 6.8 calculate the radar-echo time delay and show that one obtains minus one-half the general relativity result. Why is the sign different from the relativistic result?

6.10 Consider a Newtonian corpuscule projected radially outward from a classical point-particle field; let it begin with velocity v at radius r. If the particle has exactly the velocity necessary for escape, what is the relation between v and r? If the initial velocity is c, what is r? (This radius may be interpreted as the critical size of a Newtonian star from which light could not escape; it was discussed by Laplace in the late eighteenth century!)

6.11 Verify the theoretical values of the perihelia shifts in Table 6.1, in particular that of Icarus.

6.12 Discuss radar time delay along a radial path, e.g., between earth and Mercury at inferior conjunction.

Problems

6.1 Consider the functions λ and ν in Eq. (6.9) to be functions of both r and t. Evaluate the necessary Christoffel symbols, and show that

$$R_{01} = \frac{\partial\lambda/\partial t}{cr}.$$

Then show that the Einstein equations give $\lambda(r) + \nu(r,t) = k(t)$. Show that a change of the time coordinate similar to that in the text leads to the usual Schwarzschild solution. This is known as *Birkhoff's theorem.* It shows that the demand of time-independence in the Schwarzschild solution is superfluous (Birkhoff, 1923).

6.2 Modify the Schwarzschild solution so that $g_{00} = 1 - 2m/r$ and $g_{11} = (1 - 2\bar{m}/r)^{-1}$, where m and $\bar{m}$ may be different. Repeat the calculation of the deflection of starlight leading to (6.152) and show that the deflection is proportional to $\bar{m} + m$. That is, g_{00} and g_{11} contribute one-half each to the net deflection.

6.3 Consider a metric of the general spherically symmetric form (6.9) with the coefficients written as power series in m/r, a small quantity for celestial mechanics,

$$e^{\nu} = 1 + \alpha \frac{2m}{r} + \beta \frac{2m^2}{r^2} + \cdots$$

$$e_{\lambda} = 1 + \gamma \frac{2m}{r} + \delta \frac{2m^2}{r^2} + \cdots$$

What are the values of α, β, γ, etc., for the Schwarzschild metric? To see which terms of the Schwarzschild metric are actually tested by observation repeat the calculation of the four tests discussed in the text using this metric (Schiff, 1967).

6.4 Beginning at (6.192), introduce new dependent variables $u = H + G$, $v = H - G$ and new independent variables $y = \xi + x^0$, $z = \xi - x^0$. Show that H must be a function of y (or z) alone and G must be a function of z (or y) alone; thereby rederive Eq. (6.196).

6.5 Study the radial fall of a massive test body in Kruskal coordinates, using the standard Euler-Lagrange approach. Compare to the fall in Schwarzschild coordinates, and sketch the trajectory in the uv plane. How does a typical trajectory compare with a light-ray trajectory?

6.6 As introduced in the text, the concept of static metric involves a particular choice of coordinates. By considering a family of spacelike hypersurfaces and timelike curves orthogonal to them (hypersurface orthogonal) give a geometric and invariant characterization of a static geometry. Contrast with a geometry which is merely stationary, as discussed in Chap. 3 (see Vishveshwara, 1968, and Sec. 8.6).

6.7 In the text we studied the effect of a quadrupole moment on equatorial orbits and perihelion shifts. Study the effect on nonequatorial orbits and show that the relativistic effect and quadrupole effect are readily separated as a function of the angle between the orbital plane and the equatorial plane.

Bibliography

Ashbrook, J. (1967): What Is the True Shape of the Sun?, *Sky and Telescope*, **34**:229.

Bertotti, B., D. Brill, and R. Krotkov (1962): Experiments on Gravitation, in L. Witten (ed.), "Gravitation: An Introduction to Current Research," New York, pp. 1–45. Extensive bibliography.

Birkhoff, G. (1923): "Relativity and Modern Physics," Cambridge, Mass.

Dicke, R. H. (1964): The Sun's Rotation and Relativity, Nature, **202**:432.

Dicke, R. H., and H. M. Goldenberg (1967): Solar Oblateness and General Relativity, *Phys. Rev. Letters*, **18**:313.

Duncombe, R. L. (1956): Relativity Effects for the 3 Inner Planets, *Astron. J.*, **61**: 174–175.

Einstein, A. (1919): Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle? *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 433–436. Reprinted in Lorentz-Einstein-Minkowski: "Das Relativitätsprinzip," Leipzig (1922). English translation, "The Principle of Relativity," London (1923); reprinted by Dover, New York.

Finlay-Freundlich, E. (1955): On the Empirical Foundation of the General Theory of Relativity, in A. Beer (ed.), "Vistas in Astronomy," vol. 1, Pergamon Press, London, pp. 239–246.

Fuller, R., and J. Wheeler (1962): Causality and Multiply Connected Space-Time, *Phys. Rev.*, pp. 919–929.

Hill, H. (1974): unpublished talk. *5th Camb. Conf. Relativity*. Quoted in "Maybe the Sun Is Round After All," by B. G. Levi (1974), *Physics Today*, **27**(9), 17–19.

Klüber, H. von (1960): The Determination of Einstein's Light-deflection in the Gravitational Field of the Sun, in A. Beer (ed.), "Vistas in Astronomy," vol. 3, Pergamon Press, London, pp. 47–77.

Kruskal, M. D. (1960): Maximal Extension of Schwarzschild Metric, *Phys. Rev.*, **119**:1743.

Lieske, J. H., and G. W. Null (1969): Icarus and the Determination of Astronomical Constants, *Astron. J.*, **74**:297.

Misner, C. W., K. S. Thorne, and J. A. Wheeler (1973): "Gravitation," San Francisco, chap. 22 for gravitational collapse.

Muhleman, D. O., R. D. Ekers, and E. B. Fomalont (1970): Radio Interferometric Test of the General Relativistic Light Bending near the Sun, *Phys. Rev. Letters*, **21**:1377.

Nordtvedt, K. L. (1972): Gravitation Theory: Empirical Status from Solar System Experiments, *Science*, **178**:1157.

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

Oppenheimer, J. R., and G. M. Volkoff (1939): On Massive Neutron Cores, *Phys. Rev.*, **55**:374.

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

Schiff, L. I. (1960): On Experimental Tests of the General Theory of Relativity, *Am. J. Phys.*, **28**:340–343.

Schiff, L. I. (1967): Comparison of Theory and Observation in General Relativity, in J. Ehlers (ed.), "Relativity Theory and Astrophysics, I: Relativity and Cosmology," vol. 8 of Lectures in Applied Mathematics, Providence.

Schwarzschild, K. (1916): Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 189–196.

Seielstad, G. A., R. A. Sramek, and K. W. Weiler (1970): Measurement of the Deflection of 9.602 GHz Radiation from 3C279 in the Solar Gravitational Field, *Phys. Rev. Letters*, **21**:1373.

Shapiro, I. I. (1972): Fourth Test of General Relativity: New Radar Result, *Phys. Rev. Letters*, **26**:1132.

Shapiro, I. I., M. E. Ash, and W. B. Smith (1968): Icarus: Further Confirmation of the Relativistic Perihelion Shift, *Phys. Rev. Letters*, **20**:1517.
Shapiro, I. I., G. H. Pettengill, M. E. Ash, R. P. Ingalls, D. B. Campbell, and R. B. Dyce (1972): Mercury's Perihelion Advance: Determination by Radar, *Phys. Rev. Letters*, **28**:1594.
Shapiro, I. I., W. B. Smith, M. E. Ash, and S. Herrick (1971): General Relativity and the Orbit of Icarus, *Astron. J.*, **76**:588.
Sramek, R. A. (1971): A Measurement of the Gravitational Deflection of Microwave Radiation near the Sun, 1970 October, *Astrophys. J.*, **167**:L55.
Trumpler, R. J. (1956): Observational Results on the Light Deflection and on Red Shift in Star Spectra, in A. Mercier and M. Kervaire (eds.), "Jubilee of Relativity Theory," Basel.
Vishveshwara, C. V. (1968): Generalization of the "Schwarzschild Surface" to Arbitrary Static and Stationary Metrics, *J. Math Phys.*, **9**:1319.
Weinberg, S. (1972): "Gravitation and Cosmology," New York, chap. 11 for gravitational collapse.

CHAPTER

7

The Kerr Solution

The Schwarzschild solution, which we studied in the last chapter, describes the gravitational field exterior to a spherically symmetric body. It was obtained in 1916 and has long been a very useful model for problems in celestial mechanics and astrophysics. The problem of the field of a rotating spherical body was treated with perturbation theory by Lense and Thirring in 1918, but due to the complexity of the gravitational field equations no analogous exact solution was found for many years. In 1963 R. P. Kerr succeeded in obtaining an exact solution that represents the field exterior to a rotating axially symmetric body. The Kerr solution possesses interesting features in the region of very strong fields which are not present in the approximate Lense-Thirring solution and is therefore of particular interest in the study of the gravitational collapse of rotating stars.

In this chapter we shall derive the Kerr solution in a way that generalizes the Schwarzschild solution and discuss its physical features. We start by developing the algebraic groundwork necessary to simplify the field equations for the special metric form we shall use.

7.1 Eddington's Form of the Schwarzschild Solution

A. S. Eddington, in 1924, obtained a useful form of the Schwarzschild solution, upon which we shall base our derivation of the Kerr solution. The Schwarzschild solution is put into Eddington's form by a coordinate transformation containing a new time marker

$$(7.1)\quad \bar{x}^0 = x^0 + 2m \log\left|\frac{r}{2m} - 1\right| \qquad \bar{r} = r \qquad \bar{\theta} = \theta \qquad \bar{\varphi} = \varphi$$

By a straightforward transformation of the metric tensor we find that the Schwarzschild line element with a barred time coordinate is

$$(7.2)\quad ds^2 = (d\bar{x}^0)^2 - (dr)^2 - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2) - \frac{2m}{r}(d\bar{x}^0 + dr)^2$$

This, Eddington's form, is a flat-space line element plus a term with interesting properties. In Cartesian coordinates it is

$$(7.3)\quad \begin{aligned} ds^2 &= (d\bar{x}^0)^2 - (d\mathbf{x})^2 - \frac{2m}{r}\left(d\bar{x}^0 + \frac{x\,dx + y\,dy + z\,dz}{r}\right)^2 \\ (d\mathbf{x})^2 &= dx^2 + dy^2 + dz^2 \qquad r^2 = x^2 + y^2 + z^2 \end{aligned}$$

Thus the metric in Cartesian coordinates is

$$(7.4a)\quad g_{\mu\nu} = \eta_{\mu\nu} - 2ml_\mu l_\nu$$

$$(7.4b)\quad l_\mu = \frac{1}{\sqrt{r}}\left(1, \frac{x}{r}, \frac{y}{r}, \frac{z}{r}\right)$$

where $\eta_{\mu\nu}$ is the Lorentz metric and $l_\mu l_\nu \eta^{\mu\nu} = 0$.

The point of this exercise is the consideration of the *algebraic form* of the metric in (7.4*a*). We shall refer to this as a *degenerate* metric, for reasons to be discussed in the following sections.

7.2 Einstein's Equations for Degenerate Metrics

The preceding section serves as motivation for considering metrics of the general form

$$(7.5)\quad g_{\mu\nu} = \eta_{\mu\nu} - 2ml_\mu l_\nu \qquad l_\mu l_\nu \eta^{\mu\nu} = 0 \qquad m = \text{arbitrary constant}$$

It can be shown (see Prob. 7.1) that a metric of this form does not correspond to a Petrov type I Riemann tensor; we refer to such a space as *algebraically special* or *degenerate.* For brevity we therefore call the metric form (7.5) *degenerate;* it should be stressed, however, that not all algebraically special Riemann tensors correspond to a metric of the form (7.5) (see Prob. 7.7). As we shall see in this section and the next, the degenerate metric form greatly facilitates the purely algebraic simplification and solution of the Einstein equations.

Degenerate metrics have very interesting properties. Define a four-component upper-indexed object l^α by

$$(7.6)\quad l^\alpha \equiv \eta^{\alpha\tau} l_\tau$$

It is then easy to show that the matrix

$$(7.7)\quad g^{\mu\nu} = \eta^{\mu\nu} + 2ml^\mu l^\nu$$

is the inverse of $g_{\mu\nu}$ and therefore is the contravariant metric tensor. From this it follows that

$$(7.8)\quad l^\alpha = g^{\alpha\tau} l_\tau = \eta^{\alpha\tau} l_\tau$$

so that l^α is actually the contravariant four-vector corresponding to l_μ; its indices may be raised and lowered with either the true metric or the Lorentz metric. The vector l_μ has other interesting properties; since it is null by assumption,

$$(7.9)\quad l^\alpha l_{\alpha|\tau} = l_\alpha l^\alpha{}_{|\tau} = \tfrac{1}{2}(\eta^{\mu\nu} l_\mu l_\nu)_{|\tau} = 0$$

By enumeration it is easy to show that

$$(7.10)\quad \begin{Bmatrix} \alpha \\ \beta\ \mu \end{Bmatrix} l^\mu = -ml^\nu (l^\alpha l_\beta)_{|\nu}$$

from which we see that the covariant form of (7.9) also holds

$$(7.11)\quad l_\nu l^\nu{}_{\|\alpha} = l^\nu l_{\nu\|\alpha} = l^\nu \left(l_{\nu|\lambda} - \begin{Bmatrix} \tau \\ \nu\ \lambda \end{Bmatrix} l_\tau\right) = l^\nu l_{\nu|\lambda} = 0$$

The Einstein field equations are also simplified by our choice of metric. Consider first the metric determinant g. At any point l_μ is a flat-space null vector; that is, $l_\mu l_\nu \eta^{\mu\nu} = 0$. We can perform a proper rotation of coordinates in three-space that leaves $\eta^{\mu\nu}$ invariant and brings l_μ into the form $(a,a,0,0)$. In this system we have

$$(7.12)\quad g = \begin{Vmatrix} 1 - 2ma^2 & -2ma^2 & 0 & 0 \\ -2ma^2 & -1 - 2ma^2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{Vmatrix} = -1$$

But from the discussion in Sec. 3.5 we know that g will behave like a scalar under a three-dimensional rotation, since the Jacobian of such a transformation is 1. Thus, $g = -1$ for any degenerate metric. From this fact and Eq. (3.11) it follows that

$$\left\{ {\alpha \atop \rho\ \alpha} \right\} = \frac{\partial}{\partial x^\rho} \log \sqrt{-g} = 0 \tag{7.13}$$

Thus the field equations (5.119) are simple and contain only two terms

$$R_{\mu\nu} = -\left\{ {\alpha \atop \mu\ \nu} \right\}_{|\alpha} + \left\{ {\alpha \atop \beta\ \mu} \right\} \left\{ {\beta \atop \alpha\ \nu} \right\} = 0 \tag{7.14}$$

Because $g_{\mu\nu}$ is a polynomial in m, so is the Ricci tensor: $g_{\mu\nu}$ is first order and $R_{\mu\nu}$ is fourth order. Moreover, m is *arbitrary*, so that $g_{\mu\nu}$ must be a solution for *any* value of m. Thus, in the expression for $R_{\mu\nu}$ as a polynomial in m, each order must vanish separately. As a result, we have four sets of 10 equations to solve, which oddly enough will simplify our task. If we note that the Christoffel symbol of the first kind, $[\alpha\beta,\gamma]$, is linear in m, power counting is easy and the four sets of equations are

$$\eta^{\alpha\rho}[\mu\nu,\rho]_{|\alpha} = 0 \qquad O(m) \tag{7.15a}$$

$$2m(l^\alpha l^\rho[\mu\nu,\rho])_{|\alpha} - \eta^{\alpha\sigma}\eta^{\beta\lambda}[\beta\mu,\sigma][\alpha\nu,\lambda] = 0 \qquad O(m^2) \tag{7.15b}$$

$$l^\beta l^\lambda \eta^{\alpha\sigma}[\beta\mu,\sigma][\alpha\nu,\lambda] + l^\alpha l^\lambda \eta^{\beta\sigma}[\beta\mu,\lambda][\alpha\nu,\sigma] = 0 \qquad O(m^3) \tag{7.15c}$$

$$l^\alpha l^\sigma l^\beta l^\lambda[\beta\mu,\sigma][\alpha\nu,\lambda] = 0 \qquad O(m^4) \tag{7.15d}$$

It is easy to show by explicitly writing out terms that the order m^4 equations (7.15*d*) are satisfied identically. Similarly the order m^3 equations lead to

$$\begin{aligned} -ml_\mu l_\nu(v^\alpha v_\alpha) &= 0 \\ v^\alpha \equiv l^\beta l^\alpha{}_{\|\beta} &= l^\beta l^\alpha{}_{|\beta} \end{aligned} \tag{7.16}$$

Thus v^α is a null vector. Moreover, it is orthogonal to the null vector l^α, as is easily seen:

$$v^\nu l_\nu = (l^\alpha l^\nu{}_{|\alpha})l_\nu = l^\alpha(l_\nu l^\nu{}_{|\alpha}) = 0 \tag{7.17}$$

From this it is easy to show that v^α and l^α must be proportional. To do this we first note that the indices of v^α may be raised and lowered with the Lorentz metric, as with l^α. At any chosen point l^μ and v^μ may be written ($\mathbf{l}$ and $\mathbf{v}$ are ordinary three-vectors in Euclidean space)

$$l^\nu = (|\mathbf{l}|,\mathbf{l}) \qquad v^\nu = (|\mathbf{v}|,\mathbf{v}) \tag{7.18}$$

since they are null with respect to $\eta_{\mu\nu}$. Because they are orthogonal,

$$l^\nu v_\nu = |\mathbf{l}||\mathbf{v}|(1 - \cos\theta) = 0 \qquad \cos\theta = \frac{\mathbf{l}\cdot\mathbf{v}}{|\mathbf{l}|\cdot|\mathbf{v}|} \tag{7.19}$$

Thus $\cos\theta = 1$, and $\mathbf{v}$ is parallel to $\mathbf{l}$ at any given point. We may therefore write

$$v^\nu = l^\alpha l^\nu{}_{|\alpha} = -A(x^\mu)l^\nu \tag{7.20}$$

where A is a scalar field.

We shall defer discussion of the order m^2 equations till later, when we shall show that they are identically satisfied.

The order m equations can be further simplified. From (7.15*a*) they are

$$\eta^{\alpha\rho}[(l_\nu l_\rho)_{|\mu|\alpha} + (l_\rho l_\mu)_{|\nu|\alpha} - (l_\mu l_\nu)_{|\rho|\alpha}] = 0 \tag{7.21}$$

If we introduce the D'Alembertian operator, $\Box^2 = (\partial^2/\partial x^{0^2}) - \nabla^2$, and a scalar L defined by [we use (7.13)]

$$L = -l^\alpha{}_{\|\alpha} = -\left(l^\alpha{}_{|\alpha} + \left\{ {\alpha \atop \alpha\ \tau} \right\} l^\tau\right) = -l^\alpha{}_{|\alpha} \tag{7.22}$$

these simplify to

$$-\Box^2(l_\mu l_\nu) = [(L + A)l_\mu]_{|\nu} + [(L + A)l_\nu]_{|\mu} \tag{7.23}$$

We have so far not assumed any symmetry of the metric. In Sec. 7.4 we shall specialize to the case where the metric is stationary, or independent of x^0. A considerable simplification of the order m equations (7.23) will result.

7.3 The Order m^2 Equations

In this brief section we shall demonstrate that the order m^2 equations (7.15*b*) reduce to a scalar equation which is automatically satisfied by any solution of the order m equations (7.23). It is slightly tedious but elementary to write out all the terms indicated in (7.15b) and use (7.5), (7.9), (7.16), and (7.20) to simplify. The result is

$$(7.24)\qquad l_\mu l_\nu[2(l^\alpha A)_{|\alpha} - A^2 + l^\alpha{}_{|\beta}l^\beta{}_{|\alpha} - l^\alpha{}_{|\beta}l_\alpha{}^{|\beta}] = 0 \qquad l_\alpha{}^{|\beta} \equiv \eta^{\beta\tau}l_{\alpha|\tau}$$

This implies the scalar equation

$$(7.25)\qquad 2(l^\alpha A)_{|\alpha} - A^2 + l^\alpha{}_{|\beta}l^\beta{}_{|\alpha} - l^\alpha{}_{|\beta}l_\alpha{}^{|\beta} = 0$$

We first simplify the third term, using (7.20) and (7.22):

$$(7.26)\qquad l^\alpha{}_{|\beta}l^\beta{}_{|\alpha} = (l^\alpha{}_{|\beta}l^\beta)_{|\alpha} - l^\alpha{}_{|\beta|\alpha}l^\beta = -(Al^\alpha)_{|\alpha} + L_{|\beta}l^\beta = [(L - A)l^\alpha]_{|\alpha} + L^2$$

Similarly the fourth term can be simplified using (7.9):

$$(7.27)\qquad l^\alpha{}_{|\beta}l_\alpha{}^{|\beta} = (l^\alpha l_\alpha{}^{|\beta})_{|\beta} - l^\alpha l_\alpha{}^{|\beta}{}_{|\beta} = -l^\alpha l_\alpha{}^{|\beta}{}_{|\beta}$$

Manipulation of the order m equations will give a simpler form yet for this expression; expanding (7.23) gives

$$(7.28)\qquad -l_\mu l_\nu{}^{|\beta}{}_{|\beta} - l_\nu l_\mu{}^{|\beta}{}_{|\beta} - 2l_\mu{}^{|\beta}l_{\nu|\beta} = (L + A)_{|\mu}l_\nu + (L + A)_{|\nu}l_\mu + (L + A)(l_{\mu|\nu} + l_{\nu|\mu})$$

When contracted with l^μ, this implies

$$(7.29)\qquad -l_\nu l^\mu l_\mu{}^{|\beta}{}_{|\beta} = (L + A)_{|\mu}l^\mu l_\nu + (L + A)l_{\nu|\mu}l^\mu = l_\nu l^\mu(L + A)_{|\mu} - (L + A)Al_\nu$$

The common factor l_ν cancels, and we get

$$(7.30)\qquad \begin{aligned} -l^\mu l_\mu{}^{|\beta}{}_{|\beta} &= l^\mu(L + A)_{|\mu} - (L + A)A \\ &= [(L + A)l^\mu]_{|\mu} - l^\mu{}_{|\mu}(L + A) - A(L + A) \\ &= [(L + A)l^\mu]_{|\mu} + L^2 - A^2 \end{aligned}$$

and thus the fourth term of (7.25) can be written

$$(7.31)\qquad l^\alpha{}_{|\beta}l_\alpha{}^{|\beta} = L^2 - A^2 + [(L + A)l^\mu]_{|\mu}$$

We can now substitute the above results (7.26) and (7.31) into the left side of (7.25), to obtain

$$(7.32)\qquad \begin{aligned} 2(l^\alpha A)_{|\alpha} - A^2 + [(L - A)l^\mu]_{|\mu} + L^2 - (L^2 - A^2) - [(L + A)l^\mu]_{|\mu} \\ = 2(Al^\alpha)_{|\alpha} - 2(Al^\mu)_{|\mu} = 0 \end{aligned}$$

That is, Eq. (7.25) is satisfied identically. The entire content of the field equations is thus embodied in Eq. (7.23).

7.4 Field Equations for the Stationary Case

In the stationary, or time-independent, case it is possible to reduce the field equations to two simple partial differential equations for a single complex function. The simplification is achieved by elementary algebraic manipulation of the order m equations (7.23).

We begin by introducing a three-vector λ_j via the equation

$$(7.33)\qquad l_\mu = l_0(1,\lambda_1,\lambda_2,\lambda_3)$$

Since l_μ is a flat-space null vector ($l_\mu l_\nu \eta^{\mu\nu} = 0$), λ_j is a flat-space unit vector, $\boldsymbol{\lambda}^2 = 1$. Expressed in terms of λ_j, the order m equations are

$$(7.34a)\qquad \nabla^2(l_0^2) = 0 \qquad \mu = \nu = 0$$

$$(7.34b)\qquad \nabla^2(l_0^2\lambda_j) = [(L + A)l_0]_{|j} \qquad \mu = 0 \qquad \nu = j \neq 0$$

$$(7.34c)\qquad \nabla^2(l_0^2\lambda_i\lambda_j) = [(L + A)l_0\lambda_i]_{|j} + [(L + A)l_0\lambda_j]_{|i} \qquad \mu = i \neq 0 \qquad \nu = j \neq 0$$

and our task is to obtain l_0 and λ_j. We can manipulate these equations to obtain a first-order differential equation to replace (7.34c), which represents a considerable simplification. To do this we expand (7.34c), using (7.34a) to discard terms; (7.34b) then allows us to cancel all the second-order terms. This leaves

$$(7.35)\qquad \lambda_{i|j} + \lambda_{j|i} = \frac{2l_0}{L + A}\lambda_{i|k}\lambda_{j|k} \equiv \frac{1}{p}\lambda_{i|k}\lambda_{j|k}$$

In *this* section we shall sum over any repeated index, regardless of position; for example, k is to be summed over. The gravitational field is now described by equations (7.34a), (7.34b), and (7.35).

We shall solve (7.35) for $\lambda_{i|j}$ as a function of λ_k, with only one arbitrary parameter. To do this we denote the 3×3 matrix $\lambda_{i|j}$ by M, so that (7.35) may be written in matrix notation as

$$M + M^T = \frac{1}{p} M M^T \tag{7.36}$$

The constant length of λ_j, $\boldsymbol{\lambda}^2 = 1$, leads to an important matrix equation

$$\tfrac{1}{2}(\lambda_i \lambda_i)_{|k} = \lambda_i \lambda_{i|k} = 0 \qquad M^T \boldsymbol{\lambda} = 0 \tag{7.37}$$

That is, $\boldsymbol{\lambda}$ is in the null space of M^T. Moreover, (7.20) with $\nu = 0$ gives

$$\lambda_i l_{0|i} = A \tag{7.38}$$

and with $\nu = k \neq 0$

$$l_0 \lambda_j \lambda_{k|j} + \lambda_k \lambda_j l_{0|j} = A \lambda_k \tag{7.39}$$

Combining these, we arrive at

$$\lambda_{k|j} \lambda_j = 0 \qquad M \boldsymbol{\lambda} = 0 \tag{7.40}$$

so $\boldsymbol{\lambda}$ is *also* in the null space of M. Using only matrix algebra, we now proceed to solve (7.36), (7.37), and (7.40) for M as a function of $\boldsymbol{\lambda}$.

The nonlinear equation (7.36) may be reduced in dimension with the aid of (7.37) and (7.40). At any given point choose a 3×3 rotation matrix R that brings $\boldsymbol{\lambda}$ onto the x axis. That is,

$$R\boldsymbol{\lambda} = \boldsymbol{\lambda}' \qquad \boldsymbol{\lambda}' = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{7.41}$$

If $\boldsymbol{\lambda}$ is in the null space of M and M^T, then $\boldsymbol{\lambda}'$ is in the null space of M' and M'^T, where

$$M' = RMR^T \qquad M'^T = RM^TR^T \tag{7.42}$$

From the explicit form of $\boldsymbol{\lambda}'$ in (7.41) and the fact that $\boldsymbol{\lambda}'$ is in the null space of M' and M'^T we see that M' must have the form

$$M' = \left(\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{N'}} \\ 0 & & \end{array}\right) \qquad N' \text{ is } 2 \times 2 \tag{7.43}$$

Moreover, since matrix algebra is invariant under rotations, the matrices M' and N' may be easily seen to also obey (7.36)

$$N' + N'^T = \frac{1}{p} N'N'^T \tag{7.44}$$

This is a well-known condition related to unitary matrices; one can easily see that it is equivalent to the statement that the real matrix $I - (1/p)N'$ is unitary, i.e.,

$$U = I - \frac{1}{p} N' \qquad UU^T = U^TU = I \tag{7.45}$$

Since N' is 2×2 and real, it is therefore either a proper rotation in two dimensions or an improper rotation, i.e., a rotation plus inversion. Thus it may be written

$$\begin{gathered} U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{pmatrix} \\ |U| = \begin{cases} 1 & \text{proper case} \\ -1 & \text{improper case} \end{cases} \end{gathered} \tag{7.46}$$

We shall work only with the proper rotation matrix; our justification lies in the interesting results that follow from it. We now see from (7.45) and (7.43) that N' and M' are

$$\begin{gathered} N' = p \begin{pmatrix} 1 - \cos\theta & \sin\theta \\ -\sin\theta & 1 - \cos\theta \end{pmatrix} \\ M' = p \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 - \cos\theta & \sin\theta \\ 0 & -\sin\theta & 1 - \cos\theta \end{pmatrix} \end{gathered} \tag{7.47}$$

We must now rotate back to the original coordinates to get $M = R^TM'R$. The simple form assumed by M' allows us to write

$$\begin{aligned} M_{ik} &= R_{li}R_{jk}M'_{lj} \\ &= p(1 - \cos\theta)(R_{2i}R_{2k} + R_{3i}R_{3k}) \\ &\qquad + p\sin\theta(R_{2i}R_{3k} - R_{3i}R_{2k}) \end{aligned} \tag{7.48}$$

But since R is a rotation matrix with orthonormal rows and columns, this condenses. The orthonormality of the columns and the fact that the columns form a right-handed triad of vectors imply

$$
\begin{aligned}
R_{1i}R_{1k} + R_{2i}R_{2k} + R_{3i}R_{3k} &= \delta_{ik} \\
R_{2i}R_{3k} - R_{3i}R_{2k} &= \epsilon_{ikl}R_{1l}
\end{aligned}
\tag{7.49}
$$

Thus M_{ik} may be expressed in terms of θ and only the first row of R, $R_{1i} \equiv R_i$:

$$M_{ik} = p(1 - \cos\theta)(\delta_{ik} - R_iR_k) + p\sin\theta\,\epsilon_{ikl}R_l \tag{7.50}$$

To complete our construction of M in terms of $\boldsymbol{\lambda}$ and θ we recall from (7.41) that the matrix R rotates $\boldsymbol{\lambda}$ into $\boldsymbol{\lambda}'$, $\boldsymbol{\lambda}' = (1,0,0)$. Thus we must have, by expressing the first element of (7.41) in vector notation,

$$\mathbf{R}\cdot\boldsymbol{\lambda} = 1 \tag{7.51}$$

But $\mathbf{R}$ and $\boldsymbol{\lambda}$ are both unit vectors, and so the angle α between them must be zero,

$$\mathbf{R}\cdot\boldsymbol{\lambda} = \cos\alpha = 1 \qquad \alpha = 0 \tag{7.52}$$

and hence $\mathbf{R} = \boldsymbol{\lambda}$ or $R_i = \lambda_i$. Then we have completed our task and may express M as

$$M_{ik} = \lambda_{i|k} = p(1 - \cos\theta)(\delta_{ik} - \lambda_i\lambda_k) + p\sin\theta\,\epsilon_{ikl}\lambda_l \tag{7.53}$$

This is a most useful result; it replaces the nonlinear implicit relation (7.35) by a simple explicit expression for $\lambda_{i|k}$. Our further development rests very heavily on this equation.

A consideration of the algebraic content of (7.53) will lead to the two simple partial differential equations mentioned at the beginning of this section. We first rewrite (7.53) with parameters α and β replacing p and θ:

$$\lambda_{i|k} = \alpha(\delta_{ik} - \lambda_i\lambda_k) + \beta\epsilon_{ikl}\lambda_l \tag{7.54}$$

It will turn out that α and β determine the metric in a very elegant manner. A number of simple and important three-vector relations follow directly from (7.54). Setting $i = k$ and summing gives

$$\nabla\cdot\boldsymbol{\lambda} = 2\alpha \tag{7.55}$$

Multiplying by ϵ_{jki} and summing over i and k gives

$$\nabla\times\boldsymbol{\lambda} = -2\beta\boldsymbol{\lambda} \tag{7.56}$$

The Laplacian of $\boldsymbol{\lambda}$ may be obtained in two ways. Differentiating (7.54) with respect to x^k gives

$$\nabla^2\boldsymbol{\lambda} = \nabla\alpha - \boldsymbol{\lambda}(\nabla\alpha\cdot\boldsymbol{\lambda}) - 2(\alpha^2 + \beta^2)\boldsymbol{\lambda} + \nabla\beta\times\boldsymbol{\lambda} \tag{7.57}$$

Alternatively, from (7.56) and a well-known vector identity

$$\nabla\times(\nabla\times\boldsymbol{\lambda}) = \nabla(\nabla\cdot\boldsymbol{\lambda}) - \nabla^2\boldsymbol{\lambda} = -2(\nabla\times\beta\boldsymbol{\lambda}) \tag{7.58}$$

Solving for $\nabla^2\boldsymbol{\lambda}$ and using (7.55) and (7.56) to simplify, we then have a second form

$$\nabla^2\boldsymbol{\lambda} = 2\nabla\alpha + 2(\nabla\beta\times\boldsymbol{\lambda}) - 4\beta^2\boldsymbol{\lambda} \tag{7.59}$$

Equating the expressions (7.57) and (7.59) for $\nabla^2\boldsymbol{\lambda}$, we obtain

$$\nabla\alpha = -\nabla\beta\times\boldsymbol{\lambda} - \boldsymbol{\lambda}(\nabla\alpha\cdot\boldsymbol{\lambda}) - 2(\alpha^2 - \beta^2)\boldsymbol{\lambda} \tag{7.60}$$

This leads finally to the following very important equations:

$$\nabla\alpha\cdot\boldsymbol{\lambda} = \beta^2 - \alpha^2 \tag{7.61a}$$

$$\nabla\alpha = (\beta^2 - \alpha^2)\boldsymbol{\lambda} - \nabla\beta\times\boldsymbol{\lambda} \tag{7.61b}$$

Equations analogous to (7.61) with β replacing α on the left-hand side can be obtained. From (7.56) the divergence of $\beta\boldsymbol{\lambda}$ is zero, so that

$$\nabla\cdot\beta\boldsymbol{\lambda} = \beta(\nabla\cdot\boldsymbol{\lambda}) + \nabla\beta\cdot\boldsymbol{\lambda} \tag{7.62}$$

and thus, from (7.55),

$$\nabla\beta\cdot\boldsymbol{\lambda} = -2\alpha\beta \tag{7.63}$$

analogous to (7.61*a*). Now we cross $\boldsymbol{\lambda}$ with (7.61*b*), simplify, and solve for $\nabla\beta$:

$$\nabla\beta = \boldsymbol{\lambda}(\boldsymbol{\lambda}\cdot\nabla\beta) + (\nabla\alpha\times\boldsymbol{\lambda}) \tag{7.64}$$

or, with the use of (7.63),

$$\nabla\beta = -2\alpha\beta\boldsymbol{\lambda} + (\nabla\alpha \times \boldsymbol{\lambda}) \tag{7.65}$$

analogous to (7.61*b*).

Equations (7.61), (7.63), and (7.65) are very important and may be expressed in more concise fashion by introducing a complex function $\gamma = \alpha + i\beta$:

$$\nabla\gamma \cdot \boldsymbol{\lambda} = -\gamma^2 \tag{7.66a}$$

$$\nabla\gamma = -\gamma^2\boldsymbol{\lambda} + i(\nabla\gamma \times \boldsymbol{\lambda}) \tag{7.66b}$$

The importance of the introduction of γ should be stressed. It allows us to simplify the Einstein field equations greatly and will make the relation between the Kerr and Schwarzschild solutions transparent.

It remains in this section to obtain a pair of simple differential equations that determine γ and to show that γ in turn determines the metric, i.e., the functions l_0 and λ_j. The first differential equation is obtained by forming the Laplacian of γ from (7.66*b*) and using (7.55) and (7.56) to simplify

$$\begin{aligned} \nabla^2\gamma &= -\gamma^2\,\nabla\cdot\boldsymbol{\lambda} - 2\gamma\,\nabla\gamma\cdot\boldsymbol{\lambda} + i\,\nabla\cdot(\nabla\gamma\times\boldsymbol{\lambda}) \\ &= -2\alpha\gamma^2 + 2\gamma^3 - i\,\nabla\gamma\cdot(\nabla\times\boldsymbol{\lambda}) \\ &= -2\alpha\gamma^2 + 2\gamma^3 - 2i\beta\gamma^2 \\ &= -2\gamma^2(\alpha + i\beta - \gamma) \\ &= 0 \end{aligned} \tag{7.67}$$

Thus γ is a complex harmonic function. The second differential equation is obtained by squaring (7.66*b*) and using (7.66*a*) to simplify

$$(\nabla\gamma)^2 = \gamma^4 - (\nabla\gamma\times\boldsymbol{\lambda})^2 = \gamma^4 - ((\nabla\gamma)^2 - \gamma^4) = \gamma^4 \tag{7.68}$$

The last two differential equations determine γ. However, for maximum clarity we shall rewrite (7.68) in terms of the more convenient variable $\omega = 1/\gamma$ and repeat (7.67):

$$\nabla^2\gamma = 0 \qquad (\nabla\omega)^2 = 1 \qquad \omega \equiv \frac{1}{\gamma} \tag{7.69}$$

These are the Laplace and eikonal equations, familiar in classical optics. They determine the function γ completely, dependent upon consistent boundary conditions. More importantly, they completely replace the field equations in the various other forms in which we have written them, since, as we shall show, the metric functions l_0 and λ_j are determined by γ.

We can easily solve (7.66) for λ in terms of ω, which is more convenient than γ for this purpose. In terms of ω (7.66) reads

$$\boldsymbol{\lambda}\cdot\nabla\omega = \boldsymbol{\lambda}\cdot\nabla\omega^* = 1 \qquad \nabla\omega = \boldsymbol{\lambda} + i(\nabla\omega\times\boldsymbol{\lambda}) \tag{7.70}$$

Thus

$$\nabla\omega\times\nabla\omega^* = -i[\nabla\omega^* + \nabla\omega] + B\boldsymbol{\lambda} \tag{7.71}$$

where B represents a function of $\boldsymbol{\lambda}$ and ω which we need not write out. We can solve (7.71) for B by dotting $\nabla\omega$ into (7.71) and using (7.70):

$$B = i[1 + \nabla\omega\cdot\nabla\omega^*] \tag{7.72}$$

Now (7.71) is easily soluble for $\boldsymbol{\lambda}$

$$\boldsymbol{\lambda} = \frac{\nabla\omega + \nabla\omega^* - i(\nabla\omega\times\nabla\omega^*)}{1 + \nabla\omega\cdot\nabla\omega^*} \tag{7.73}$$

It now remains only to obtain l_0 in terms of γ. The function l_0 must satisfy Eqs. (7.34*a*) and (7.34*b*). We shall show that

$$l_0^2 = \mathrm{Re}\,(\gamma) = \alpha \tag{7.74}$$

or a multiple of this is the unique solution of these equations consistent with (7.34*c*). From (7.69) we know that α is harmonic, so that (7.34*a*) is immediately satisfied. To show that $l_0^2 = \alpha$ is a solution of (7.34*b*) we first calculate the left-hand side using $l_0^2 = \alpha$ as a trial solution, with α harmonic

$$\nabla^2(\alpha\lambda_j) = \alpha\,\nabla^2\lambda_j + 2\alpha_{|k}\lambda_{j|k} \tag{7.75}$$

With the aid of (7.54) and (7.59) this becomes

$$\nabla^2(\alpha\lambda_j) = 4\alpha\,\nabla\alpha + 2\alpha(\alpha^2 + \beta^2)\boldsymbol{\lambda} + 2\alpha(\nabla\beta\times\boldsymbol{\lambda}) + 2\beta(\nabla\alpha\times\boldsymbol{\lambda}) \tag{7.76}$$

or, using (7.61*b*) and (7.65) to simplify further,

$$\nabla^2(\alpha\boldsymbol{\lambda}) = 2\alpha\,\nabla\alpha + 2\beta\,\nabla\beta = \nabla(\alpha^2 + \beta^2) \tag{7.77}$$

We next calculate the right-hand side of (7.34*b*). From the definitions $\alpha = p(1 - \cos\theta)$ and $\beta = p \sin\theta$ preceding (7.54) we find

$$\alpha^2 + \beta^2 = 2p^2(1 - \cos\theta) = 2\alpha p \tag{7.78}$$

From the definition of p in (7.35),

$$p = \frac{L + A}{2l_0} \tag{7.79}$$

We then have a relation between $L + A$ and α and β,

$$A + L = \frac{l_0}{\alpha}(\alpha^2 + \beta^2) \tag{7.80}$$

Thus with $l_0^2 = \alpha$ the right-hand side of (7.34*b*) is

$$[(L + A)l_0]_{|j} = (\alpha^2 + \beta^2)_{|j} \tag{7.81}$$

Comparing this with the right-hand side as obtained in (7.77), we see that $l_0^2 = \alpha$ is indeed a solution. It is moreover easily shown to be the unique solution, up to a multiplicative constant (see Exercise 7.3).

Let us now summarize the extensive simplification we have made. The field equations reduce to the pair of simple equations (7.69), with the metric functions given explicitly by (7.73) and (7.74). As we shall see in more detail, the complex function γ plays the role of a generalized Newtonian potential since it obeys Laplace's equation, and in the weak-field limit Re (γ) is precisely the Newtonian potential.

7.5 The Schwarzschild and Kerr Solutions

We now wish to solve the Einstein field equations for the stationary degenerate metric. These equations have now been greatly simplified to the form in Eq. (7.69), with (7.73) and (7.74) giving the metric functions explicitly. In analogy with Newtonian theory we first consider the simple spherically symmetric solution to Laplace's equation:

$$\gamma = \frac{1}{r} = [x^2 + y^2 + z^2]^{-1/2} \tag{7.82}$$

It is easily checked that $\omega = r$ satisfies the eikonal equation, so that (7.82) is a solution of the system (7.69). The metric function l_0^2, the analogue of the Newtonian potential, and the vector λ_i are then easily

obtained from (7.73) and (7.74)

$$l_0^2 = \frac{1}{r} \qquad \lambda_1 = \frac{x}{r} \qquad \lambda_2 = \frac{y}{r} \qquad \lambda_3 = \frac{z}{r} \tag{7.83}$$

Thus from the definitions (7.33) and (7.5) we obtain the line element

$$ds^2 = (dx^0)^2 - (d\mathbf{x})^2 - \frac{2m}{r}\left(dx^0 + \frac{x\,dx + y\,dy + z\,dz}{r}\right)^2 \tag{7.84}$$

This is precisely the Eddington form of the Schwarzschild solution discussed in Sec. 7.1. From this fact we can now identify the arbitrary parameter m as the geometric mass of the source.

We now ask for the simplest generalization of the Schwarzschild solution in the above context. We are naturally led to consider a general displacement of the origin, i.e.,

$$\gamma = \frac{1}{r} = [(x - a_1)^2 + (y - a_2)^2 + (z - a_3)^2]^{-1/2} \qquad a_i = \text{const} \tag{7.85}$$

since this satisfies (7.69) for any choice of the constants a_i. However, if the a_i are real, this solution corresponds to a physical displacement of the origin and is of no physical interest. On the other hand, an imaginary set of a_i represents a new physical situation. Without loss of generality (see Exercise 7.5) we may write such an imaginary displaced γ function as

$$\gamma = (x^2 + y^2 + (z - ia)^2)^{-1/2} \tag{7.86}$$

This represents the Kerr solution.

From this function we can obtain the metric functions l_0^2 and λ_i from (7.73) and (7.74), just as with the Schwarzschild case above, but with slightly more algebra. We first split ω into real and imaginary parts for convenience

$$\omega = \rho + i\sigma = (r^2 - a^2 - 2iaz)^{1/2} \qquad r^2 \equiv x^2 + y^2 + z^2 \tag{7.87}$$

Squaring this equation and equating real and imaginary parts, we get

$$\rho^2 - \sigma^2 = r^2 - a^2 \qquad \sigma = -\frac{az}{\rho} \tag{7.88}$$

from which we obtain a quadratic relation between ρ^2 and the Cartesian markers

$$\rho^4 - \rho^2(r^2 - a^2) - a^2z^2 = 0 \tag{7.89}$$

Explicitly ρ^2 is given by

$$\rho^2 = \frac{r^2 - a^2}{2} + \left[\frac{(r^2 - a^2)^2}{4} + a^2z^2\right]^{1/2} \tag{7.90}$$

Note that we have chosen the plus sign so that for $r \gg a$ the marker ρ is asymptotically equal to r, as is necessary according to (7.87). It is now easy to write out γ and obtain $l_0^2 = \mathrm{Re}\,(\gamma)$ in a useful and simple form

$$\begin{aligned} \gamma &= \frac{1}{\rho + i\sigma} = \frac{\rho}{\rho^2 + \sigma^2} - \frac{i\sigma}{\rho^2 + \sigma^2} \\ l_0^2 &= \frac{\rho}{\rho^2 + \sigma^2} = \frac{\rho^3}{\rho^4 + a^2z^2} \end{aligned} \tag{7.91}$$

where we have used (7.88) to eliminate σ.

To obtain the vector $\boldsymbol{\lambda}$ from (7.73) we first calculate $\nabla\omega$ from (7.87)

$$\nabla\omega = \frac{\mathbf{r} - ia\hat{\mathbf{k}}}{\omega} \tag{7.92}$$

where $\hat{\mathbf{k}}$ is the unit three-vector along the z axis, (0,0,1). It then follows by substitution into (7.73) that

$$\boldsymbol{\lambda} = \frac{2[\rho\mathbf{r} - a\sigma\hat{\mathbf{k}} + a(\mathbf{r} \times \hat{\mathbf{k}})]}{|\omega|^2 + r^2 + a^2} \tag{7.93}$$

We can simplify this by noting from (7.87) and (7.90) that

$$\begin{aligned} |\omega|^2 &= [(r^2 - a^2)^2 + 4a^2z^2]^{1/2} = 2\rho^2 - (r^2 - a^2) \\ |\omega|^2 + r^2 + a^2 &= 2(\rho^2 + a^2) \end{aligned} \tag{7.94}$$

Thus, $\boldsymbol{\lambda}$ can be written in quite simple form as

$$\boldsymbol{\lambda} = \frac{\rho}{\rho^2 + a^2}\left[\mathbf{r} + \frac{a^2z}{\rho^2}\hat{\mathbf{k}} + \frac{a}{\rho}(\mathbf{r} \times \hat{\mathbf{k}})\right] \tag{7.95}$$

or in terms of components

$$\begin{aligned} \lambda_1 &= \frac{\rho x + ay}{a^2 + \rho^2} \\ \lambda_2 &= \frac{\rho y - ax}{a^2 + \rho^2} \\ \lambda_3 &= \frac{z}{\rho} \end{aligned} \tag{7.96}$$

This may be compared with the Eddington form of the Schwarzschild solution as given in (7.83); for r much larger than a the two solutions are asymptotically equal, so that the present solution behaves like the Schwarzschild solution far from the source at the origin, as should be expected from the starting point (7.86).

Finally, let us put the pieces together as specified in (7.5) and (7.33) in order to display the Kerr line element

$$\begin{aligned} ds^2 = (dx^0)^2 - (d\mathbf{x})^2 - \frac{2m\rho}{\rho^4 + a^2z^2}\bigg[dx^0 &+ \frac{\rho}{a^2 + \rho^2}(x\,dx + y\,dy) \\ &+ \frac{a}{a^2 + \rho^2}(y\,dx - x\,dy) + \frac{z}{\rho}\,dz\bigg]^2 \end{aligned} \tag{7.97}$$

This is in the form obtained by Kerr in 1963.

7.6 Other Coordinates

The degenerate form of the metric tensor (7.5) has proved to be tremendously convenient for generalizing the Eddington form of the Schwarzschild solution to the Kerr solution, (7.97). We shall presently show that it actually represents the axially symmetric gravitational field of a rotating mass. Preparatory to showing this we wish to introduce a second set of coordinates that illustrate the axial symmetry very clearly. Then we shall introduce a third set of coordinates that is most convenient for analyzing the physical consequences.

As a radial coordinate we choose in preference to r the marker ρ as introduced in (7.87) or given explicitly in (7.90). It is then natural to introduce an angular coordinate θ as with the usual polar coordinates

$$\cos\theta = \frac{z}{\rho} \tag{7.98}$$

Next we introduce an angular coordinate φ by a convenient and elegant complex expression

$$(\rho - ia)e^{i\varphi}\sin\theta = x + iy \tag{7.99}$$

Finally we choose a time coordinate

$$u = x^0 + \rho \tag{7.100}$$

It is a simple task to express the line element (7.97) in terms of these new coordinates. The flat-space part is readily obtained from (7.98) and (7.99) by simple manipulations

$$
\begin{aligned}
(7.101) \qquad dz^2 &= (\cos\theta\, d\rho - \rho \sin\theta\, d\theta)^2 \\
dx^2 + dy^2 &= |d(x+iy)|^2 = |d(\rho - ia)e^{i\varphi} \sin\theta|^2 \\
&= (\sin\theta\, d\rho + a \sin\theta\, d\varphi + \rho\cos\theta\, d\theta)^2 \\
&\qquad + (\rho \sin\theta\, d\varphi - a\cos\theta\, d\theta)^2
\end{aligned}
$$

Similarly, several of the other differential expressions that occur in the line element (7.97) are obtained in simple steps from (7.98) and (7.99) as

$$
\begin{aligned}
(7.102a) \qquad x\,dx + y\,dy &= \tfrac{1}{2} d|x+iy|^2 = \tfrac{1}{2} d[(\rho^2 + a^2)\sin^2\theta] \\
&= (\rho^2 + a^2)\sin\theta\cos\theta\, d\theta + \rho \sin^2\theta\, d\rho
\end{aligned}
$$

$$
\begin{aligned}
(7.102b) \qquad x\,dy - y\,dx &= -\mathrm{Im}\,[(x+iy)d(x-iy)] \\
&= -\mathrm{Im}\,\{[(\rho - ia)e^{i\varphi}\sin\theta] d[(\rho + ia)e^{-i\varphi}\sin\theta]\} \\
&= (\rho^2 + a^2)\sin^2\theta\, d\varphi + a \sin^2\theta\, d\rho
\end{aligned}
$$

$$
(7.102c) \qquad z\,dz = \rho\cos^2\theta\, d\rho - \rho^2 \sin\theta\cos\theta\, d\theta
$$

Lastly, we note from (7.98) and (7.100)

$$
(7.103) \qquad \frac{2m\rho^3}{\rho^4 + a^2 z^2} = \frac{2m\rho}{\rho^2 + a^2\cos^2\theta} \qquad dx^0 = du - d\rho
$$

If the above expressions (7.101) and (7.103) are substituted into the line element (7.97), the result is a line element in the new coordinates given by

$$
\begin{aligned}
(7.104) \qquad ds^2 = {}& \left(1 - \frac{2m\rho}{\rho^2 + a^2\cos^2\theta}\right) du^2 - (\rho^2 + a^2\cos^2\theta)\, d\theta^2 \\
&- \left[(\rho^2 + a^2)\sin^2\theta + \frac{2m\rho a^2 \sin^4\theta}{\rho^2 + a^2\cos^2\theta}\right] d\varphi^2 - 2du\,d\rho \\
&- \frac{4m\rho a \sin^2\theta}{\rho^2 + a^2\cos^2\theta}\, du\,d\varphi - 2a\sin^2\theta\, d\rho\, d\varphi
\end{aligned}
$$

There is no dependence of the line element on the angular coordinate φ, so that the solution (7.104) is manifestly axially symmetric.

A further simplification of this result can be made. We wish to show ultimately that the Kerr solution is appropriate to some form of rotating body, and so it is reasonable to suppose that space-time is, loosely speaking, dragged around with the body. This suggests that we attempt to make the line element formally as similar as possible to that of rotating flat space, (4.83). This form of the flat-space line element contains only one off-diagonal term in the metric tensor, a term in $d\varphi\, dt$. Motivated by this physical consideration, as well as mathematical simplicity, we shall introduce a coordinate transformation that eliminates the $du\, d\rho$ and $d\rho\, d\varphi$ terms of (7.104).

Let us write the line element (7.104) as

$$
\begin{aligned}
(7.105) \qquad ds^2 = {}& g_{00}\, du^2 + g_{22}\, d\theta^2 + g_{33}\, d\varphi^2 + 2g_{03}\, du\, d\varphi \\
&+ 2g_{01}\, du\, d\rho + 2g_{13}\, d\rho\, d\varphi
\end{aligned}
$$

We guess a simple form for the desired transformation

$$
(7.106) \qquad \begin{aligned}
c\hat{t} &= u - A(\rho) \qquad & du &= c\, d\hat{t} + A'\, d\rho \\
\hat{\varphi} &= \varphi - B(\rho) \qquad & d\hat{\varphi} &= d\varphi + B'\, d\rho
\end{aligned}
$$

where $A(\rho)$ and $B(\rho)$ are functions only of ρ, to be determined, and a prime denotes differentiation with respect to ρ. If du and $d\varphi$ from (7.106) are substituted into (7.105), the result is

$$
\begin{aligned}
(7.107) \qquad ds^2 = {}& g_{00}c^2\, dt^2 + (g_{00}A'^2 + g_{33}B'^2 + 2g_{01}A' + 2g_{13}B' \\
&+ 2g_{03}A'B')\, d\rho^2 + g_{22}\, d\theta^2 + g_{33}\, d\hat{\varphi}^2 + 2g_{03}c\, d\hat{t}\, d\hat{\varphi} \\
&+ 2(A'g_{03} + B'g_{33} + g_{13})\, d\hat{\varphi}\, d\rho + 2(A'g_{00} + B'g_{03} + g_{01})c\, d\hat{t}\, d\rho
\end{aligned}
$$

We now demand that the coefficients of $d\hat{\varphi}\, d\rho$ and $d\hat{t}\, d\rho$ be zero, which provides us with a simple set of linear equations for A' and B'. The solutions of these linear equations are

$$
(7.108) \qquad \begin{aligned}
A' &= \frac{g_{33}g_{01} - g_{03}g_{13}}{g_{03}^2 - g_{00}g_{33}} = \frac{\rho^2 + a^2}{\rho^2 + a^2 - 2m\rho} \\
B' &= \frac{g_{00}g_{13} - g_{03}g_{01}}{g_{03}^2 - g_{00}g_{33}} = \frac{-a}{\rho^2 + a^2 - 2m\rho}
\end{aligned}
$$

It is very important to note that A' and B' are functions of ρ only. This allows one to integrate (7.108) for $A(\rho)$ and $B(\rho)$, although we shall not do this explicitly.

The line element is now given by (7.107) with the last two terms vanishing by construction. We can shorten the rather lengthy coefficient of $d\rho^2$ in (7.107).

$$(7.109)\quad \begin{aligned} g_{00}A'^2 + g_{33}B'^2 + 2g_{01}A' + 2g_{13}B' + 2g_{03}A'B' &= A'(A'g_{00} + B'g_{03} + g_{01}) + B'(A'g_{03} + B'g_{33} + g_{13}) \\ &\qquad + g_{01}A' + g_{13}B' \\ &= g_{01}A' + g_{13}B' \end{aligned}$$

In the last step we have utilized the fact that A' and B' are chosen so that the coefficients of $d\hat{\varphi}\, d\rho$ and $d\hat{t}\, d\rho$ in (7.107) are zero.

It is now straightforward to substitute the specific metric functions from (7.104) and (7.108) into the line element (7.107), with the coefficient of $d\rho^2$ given in (7.109)

$$(7.110)\quad \begin{aligned} ds^2 = \left(1 - \frac{2m\rho}{\rho^2 + a^2\cos^2\theta}\right) c^2\, d\hat{t}^2 - \frac{\rho^2 + a^2\cos^2\theta}{\rho^2 + a^2 - 2m\rho}\, d\rho^2 \\ - (\rho^2 + a^2\cos^2\theta)\, d\theta^2 - \left[(\rho^2 + a^2)\sin^2\theta + \frac{2m\rho a^2\sin^4\theta}{\rho^2 + a^2\cos^2\theta}\right] d\hat{\varphi}^2 \\ - 2\frac{2m\rho a\sin^2\theta}{\rho^2 + a^2\cos^2\theta}\, c\, d\hat{t}\, d\hat{\varphi} \end{aligned}$$

In this form (Boyer and Lindquist, 1967) the Kerr metric is manifestly axially symmetric, closely resembles the Schwarzschild solution in its standard form (6.53), and contains only one cross term, $d\hat{t}\, d\hat{\varphi}$, in analogy with the rotating flat-space line element (4.83).

Finally, it should be stressed once more that the marker ρ is quite as good a radial marker as the marker r previously used; in terms of either coordinate the Einstein equations are satisfied.

7.7 The Kerr Solution and Rotation

In this section we wish to demonstrate that the Kerr metric represents the field exterior to an axially symmetric rotating body, give a physical interpretation of the parameter a, and discuss some physical effects associated with the rotation of the source.

We first note that the Kerr metric in the form (7.110) has the general features expected for the field of a rotating source. (1) It is axially symmetric and time-independent. (2) It is unchanged if the signs of φ

and t are both reversed. (We omit the hat in this section.) This is physically reasonable since φ and t are merely markers, and the signs are conventional: running time backward with a negative spin direction should be physically equivalent to running time forward with a positive spin direction. (3) For $a = 0$ it reduces to the Schwarzschild metric in the standard form (6.53). (4) It is unchanged if the signs of φ and a are both reversed, suggesting that a specifies a spin direction. (5) The presence of the $d\varphi\, dt$ cross term makes it superficially but suggestively similar to the metric of rotating flat space, (4.83).

The problem of physically interpreting the parameter a is made difficult by the lack of a classical analogue of the Kerr metric. In classical gravitational theory the field of an axially symmetric body is independent of its rotational motion, unlike the situation in relativity. Lacking a *direct* classical analogue, we shall instead use the results of an approximate relativistic calculation to build a bridge between classical and relativistic concepts. In 1918 Lense and Thirring studied the gravitational field of a spinning sphere of constant density. Using relativistic field equations valid in regions of space containing mass-energy (which we discuss in Chap. 10), they were able to obtain an approximate solution, valid for low rates of spin and weak fields, both inside and outside the sphere. The solution exterior to the sphere has the form (see Prob. 10.2)

$$(7.111)\quad \begin{aligned} ds^2 = \left(1 - \frac{2m}{r}\right) c^2\, dt^2 - \left(1 + \frac{2m}{r}\right) d\sigma^2 \\ + 4\frac{\kappa J}{c^3 r}\sin^2\theta\, d\varphi\, c\, dt \qquad r = [x^2 + y^2 + z^2]^{1/2} \end{aligned}$$

where m is the geometric mass, J is the angular momentum of the source, and $d\sigma^2$ is the flat-space line element in three dimensions. This solution is valid only to first order in the dimensionless quantities m/r and $\kappa J/c^3r^2$. This is a consistent expansion for rotating stars that are much larger than their Schwarzschild radius (see Exercise 7.7).

The Kerr solution (7.110) can be reduced to this approximate form in three steps. First we expand to first order in a/ρ:

$$(7.112)\quad \begin{aligned} ds^2 = \left(1 - \frac{2m}{\rho}\right) c^2\, dt^2 - \left(1 - \frac{2m}{\rho}\right)^{-1} d\rho^2 \\ - \rho^2(d\theta^2 + \sin^2\theta\, d\varphi^2) - \frac{4ma}{\rho}\sin^2\theta\, d\varphi\, c\, dt \end{aligned}$$

where to first order $\rho = r$. This is merely the exact Schwarzschild solution plus a cross term proportional to a. To compare with (7.111) we must put this into isotropic form; this is very easily done using the same change of radial coordinate as we used in Chap. 6 to put the Schwarzschild line element into isotropic form, $\rho = \hat{\rho}(1 + m/2\hat{\rho})^2$, where $\hat{\rho}$ is the isotropic radial marker. This leads to (the algebra need not be repeated; see Sec. 6.2)

$$(7.113)\quad ds^2 = \frac{(1 - m/2\hat{\rho})^2}{(1 + m/2\hat{\rho})^2} c^2\, dt^2 - \left(1 + \frac{m}{2\hat{\rho}}\right)^4 d\sigma^2 - \frac{4ma}{\hat{\rho}(1 + m/2\hat{\rho})^2} \sin^2\theta\, d\varphi\, c\, dt$$

Finally we expand to first order in $m/\hat{\rho}$

$$(7.114)\quad ds^2 = \left(1 - \frac{2m}{\hat{\rho}}\right) c^2\, dt^2 - \left(1 + \frac{2m}{\hat{\rho}}\right) d\sigma^2 - \frac{4ma}{\hat{\rho}} \sin^2\theta\, d\varphi\, c\, dt$$

which agrees with (7.111) if we identify

$$(7.115)\quad ma = -\frac{\kappa J}{c^3}$$

Thus a is a measure of the angular momentum per unit mass of the source. We shall refer to ma as the geometric angular momentum in the same spirit as we refer to m as the geometric mass. The negative sign in (7.115) should be stressed; a body rotating in a positive sense will have a positive J and a negative a.

This correspondence argument has relied on an approximate solution for a spherical rotating source. However, we need not conclude that the exact Kerr solution necessarily corresponds to a spherical source and will use only the fact that the source has a geometric angular momentum equal to ma without reference to its structure (see Prob. 11.3).

A very interesting physical effect results from the rotational nature of the Kerr solution; a body in geodesic motion experiences a force proportional to the parameter a reminiscent of a Coriolis force. Loosely speaking, we may think of the rotating source as "dragging" space around with it; in a Machian sense the source "competes" with the Lorentzian boundary conditions at infinity in the establishment of a local inertial frame. To demonstrate this force we proceed to obtain the geodesic equations of motion, precisely as in Chap. 6 [see (6.74) to (6.78)]. For simplicity we shall consider the approximate form (7.111) and work only to first order

in a/ρ. The equations of motion are

$$\frac{d}{ds}(\rho^2\dot{\theta}) = \rho^2 \sin\theta\cos\theta\,\dot{\varphi}^2 + \frac{4ma}{\rho}\sin\theta\cos\theta\, c\dot{t}\dot{\varphi}$$

$$\frac{d}{ds}\left(\rho^2\sin^2\theta\,\dot{\varphi} + \frac{2ma}{\rho}\sin^2\theta\, c\dot{t}\right) = 0$$

$$(7.116)\quad \frac{d}{ds}\left(\left(1 - \frac{2m}{\rho}\right)c\dot{t} - \frac{2ma}{\rho}\sin^2\theta\,\dot{\varphi}\right) = 0$$

$$1 = \left(1 - \frac{2m}{\rho}\right)c^2\dot{t}^2 - \left(1 - \frac{2m}{\rho}\right)^{-1}\dot{\rho}^2 - \rho^2(\dot{\theta}^2 + \sin^2\theta\,\dot{\varphi}^2) - \frac{4ma}{\rho}\sin^2\theta\, c\dot{t}\dot{\varphi}$$

We limit ourselves to the special case of equatorial orbits, $\theta = \pi/2$, which is particularly simple (note, however, that the general orbit in the Kerr metric does *not* lie in a plane, unlike the Schwarzschild case):

$$(7.117a)\quad \frac{d}{ds}\left(\rho^2\dot{\varphi} + \frac{2ma}{\rho}c\dot{t}\right) = 0$$

$$(7.117b)\quad \frac{d}{ds}\left[\left(1 - \frac{2m}{\rho}\right)c\dot{t} - \frac{2ma}{\rho}\dot{\varphi}\right] = 0$$

$$(7.117c)\quad 1 = \left(1 - \frac{2m}{\rho}\right)c^2\dot{t}^2 - \left(1 - \frac{2m}{\rho}\right)^{-1}\dot{\rho}^2 - \rho^2(\dot{\theta}^2 + \sin^2\theta\,\dot{\varphi}^2) - \frac{4ma}{\rho}\sin^2\theta\, c\dot{t}\,\dot{\varphi}$$

The conserved angular momentum conjugate to φ is not $\rho^2\dot{\varphi}$, as in the Schwarzschild problem, but the quantity indicated in (7.117a); an extra term proportional to a appears.

We now consider, as in Sec. 4.2, a body in instantaneous radial motion, $\dot{\varphi} = 0$. Equation (7.117a) then yields an equation for $\ddot{\varphi}$

$$(7.118)\quad \rho\ddot{\varphi} + \frac{2ma}{\rho^2}c\ddot{t} - \frac{2ma}{\rho^3}c\dot{t}\dot{\rho} = 0$$

while (7.117b) gives

$$(7.119)\quad c\dot{t} = l\left(1 - \frac{2m}{\rho}\right)^{-1} \qquad c\ddot{t} = \frac{2ma}{\rho(1 - 2m/\rho)}\ddot{\varphi} - \frac{2m\dot{\rho}l}{\rho^2(1 - 2m/\rho)^2}$$

We substitute these back into (7.118) and obtain, to first order in a/ρ,

$$\rho\ddot{\varphi} + 2\left[\frac{-ma}{\rho^3(1 - 2m/\rho)}\right]\left[1 + \frac{4m}{\rho(1 - 2m/\rho)}\right]\dot{\rho} = 0 \tag{7.120}$$

and to first order in m/ρ

$$\rho\ddot{\varphi} + 2\left(\frac{-ma}{\rho^3}\right)\dot{\rho} = 0 \tag{7.120$'$}$$

This is identical in form to the Coriolis equation (4.123) with a function of ρ replacing the constant angular velocity w. This Coriolislike force is of course fundamentally different from the true Coriolis force in (4.123), which can be transformed away globally by a coordinate transformation.

7.8 Distinguished Surfaces and the Rotating Black Hole

In the Kerr solution two surfaces arise which are analogous to the Schwarzschild singular surface and which are of great physical interest, e.g., in the gravitational collapse of a rotating star. We devote this section to their study.

Let us first study the red shift of light emitted from a source at rest in the Kerr geometry. We obtained a general solution to this problem in Eq. (4.152):

$$\nu = \nu_0\left(\frac{g_{00}(x_s{}^\mu)}{g_{00}(x^\mu)}\right)^{1/2} \tag{7.121}$$

Here ν is the frequency of the light observed at x^μ, and ν_0 is the proper frequency of the light emitted by the source at rest at $x_s{}^\mu$. [Note that the last term appearing in Eq. (4.152) is an approximation and not appropriate to the present discussion.] It is clear from (7.110) that for large values of ρ the red shift in the Kerr metric is approximately equal to that in the Schwarzschild metric, since the g_{00} functions differ only by terms of second order in a/ρ. However, the Kerr metric displays two surfaces of infinite red shift, both of which are intrinsically different from the Schwarzschild surface of infinite red shift. These occur where $g_{00}(x_s{}^\mu) = 0$, which from (7.121) makes $\nu = 0$. Setting g_{00} in (7.110) equal to zero, we obtain two axially symmetric surfaces

$$\rho = m \pm \sqrt{m^2 - a^2\cos^2\theta} \tag{7.122}$$

In the limit of $a \to 0$ these surfaces reduce to the Schwarzschild surface $\rho = 2m$, for the plus sign, and the origin $\rho = 0$, for the minus sign. The surface corresponding to the plus sign is of much greater physical interest; it is an axially symmetric surface with a radius at the equator of $2m$ and a radius at the poles of $m + \sqrt{m^2 - a^2}$. We shall assume that $|a| < m$ so that the surface is well defined; that is, ρ is real. The surface corresponding to the minus sign is completely contained within the above surface, $\rho_\infty \equiv m + \sqrt{m^2 - a^2\cos^2\theta}$.

Let us note that the outer infinite red shift surface is comparable in radius to the Schwarzschild surface, so that for ordinary stars like the sun it is of little physical consequence, lying well *inside* the star, where the vacuum field equations are not valid. Only for extremely dense stars can regions near this surface be in free space.

The Schwarzschild singular surface studied in Chap. 6 plays two unusual roles: (1) it serves as an infinite red shift surface for sources at rest, and (2) it acts as a one-way membrane for physical objects whose trajectories lie in or on the forward light cone, as discussed in Sec. 6.8. In the Kerr metric these roles are played by different surfaces. To find the one-way membranes of the Kerr solution we shall introduce and study the concept of a so-called null hypersurface, i.e., a hypersurface whose normal vector is null. Such a null hypersurface always acts as a one-way membrane, as we shall show.

Consider a smooth hypersurface S defined by the equation

$$u(x^\mu) = \text{const} \tag{7.123}$$

the vector $n_\alpha = u_{|\alpha}$ is a normal to S since its inner product with any dx^α contained in S is zero:

$$n_\alpha\, dx^\alpha = u_{|\alpha}\, dx^\alpha = du = 0 \tag{7.124}$$

At any point P on S we introduce a locally Lorentzian metric so that the line element is

$$ds^2 = (dx^0)^2 - (dx)^2 - (dy)^2 - (dz)^2 \tag{7.125}$$

and the local light cone, defined as the local hypersurface $ds^2 = 0$, is very simple. Moreover, by a rotation in three-space we can place the three-vector part of n^α along the x axis

$$n^\alpha = (n^0, n^1, 0, 0) \qquad n^\alpha n_\alpha = (n^0)^2 - (n^1)^2 \tag{7.126}$$

This form of the normal vector restricts the form of any vector t^α tangent to S at P since the two must be orthogonal:

$$n_\alpha t^\alpha = n^0 t^0 - n^1 t^1 = 0 \qquad \frac{t^0}{t^1} = \frac{n^1}{n^0} \tag{7.127}$$

thus t^α must have the form

$$t^\alpha = \lambda(n^1, n^0, a, b) \tag{7.128}$$

where λ, a, and b are arbitrary. It follows that the norm of t^α is

$$\begin{aligned} t^\alpha t_\alpha &= \lambda^2[(n^1)^2 - (n^0)^2 - a^2 - b^2] \\ &= -\lambda^2(n^\alpha n_\alpha + (a^2 + b^2)) \end{aligned} \tag{7.129}$$

This simple relation between the norms of the normal and tangent vectors of S leads to a beautiful geometrical result. We must consider three cases in the light of (7.129).

Case I: n^α is timelike, $n^\alpha n_\alpha > 0$. Then $t^\alpha t_\alpha$ is negative-definite so that t^α is spacelike and *no* tangent vector to S can lie on the local light cone, i.e., be null.

Case II: n^α is null, $n^\alpha n_\alpha = 0$. Then $t^\alpha t_\alpha$ is negative *except* when $a = b = 0$, in which case it is zero. Thus there is *one* unit tangent vector to S, which, along with its multiples, lies on the local light cone at P.

Case III: n^α is spacelike, $n^\alpha n_\alpha < 0$. Then $t^\alpha t_\alpha$ may have either sign or be zero. In particular $t^\alpha t_\alpha = 0$ on the circle defined by $a^2 + b^2 = -n^\alpha n_\alpha > 0$. Thus there is a *whole family* of vectors tangent to S, which also lie on the local light cone.

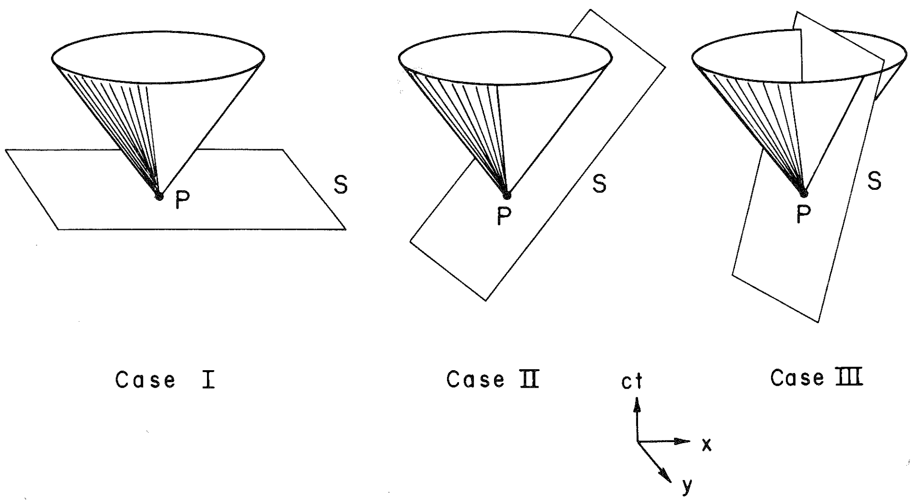


Fig. 7.1
The three possible relations of a hypersurface S to the local light cone, with one space dimension suppressed.

We may picture the above results if we suppress the z coordinate and consider only two spatial coordinates and one time coordinate. The geometry is shown in Fig. 7.1. It is clear also that the final statement of the result has an invariant meaning.

The physical interpretation of this result is easy. Since light has a trajectory on the forward light cone and massive bodies have trajectories within the forward light cone, we see that physical objects can pass through a spacelike hypersurface in *either* direction and can pass through a timelike hypersurface in *only one* direction. The null hypersurface is the critical case: it is the configuration where the one-way behavior begins, and we may therefore identify it as a *one-way membrane*. A very simple example of this behavior can be found in special relativity. The hypersurface t = const is timelike, and physical objects can pass it in only one direction; x = const is spacelike, and physical objects may pass in either direction; while $ct - x = 0$ is null and acts as a one-way membrane. In the last case, a tangent vector of the null hypersurface that lies along the local light cone is $t^\alpha = (1,1,0,0)$. A second example of a one-way membrane is provided by the Schwarzschild singular surface. A spherical surface r = const in the Schwarzschild geometry has a normal vector

$$n_\alpha = (0,1,0,0) \qquad n^\alpha n_\alpha = -\left(1 - \frac{2m}{r}\right) \tag{7.130}$$

Thus as r decreases through $2m$ the spherical surface changes from spacelike to null to timelike. Unlike the above example in special relativity, this surface has the important feature that it is finite in spatial extent.

In Chap. 8 we shall discuss null hypersurfaces further. They occur also as characteristic hypersurfaces of the Einstein equations, the hypersurfaces over which the second derivatives of the metric may be discontinuous, analogous to the characteristic surfaces of Maxwell's equations discussed in Chap. 4.

We now search for the null hypersurfaces of the Kerr geometry. It is easy to check that the outer infinite red shift surface (7.122) is not a null hypersurface; the normal vector and its norm are

$$n_\alpha = \left(0, 1, -\frac{a^2 \cos\theta \sin\theta}{\sqrt{m^2 - a^2\cos^2\theta}}, 0\right)$$
(7.131)
$$n^\alpha n_\alpha = -\frac{\rho^2 + a^2 - 2m\rho + (a^4\cos^2\theta\sin^2\theta)/(m^2 - a^2\cos^2\theta)}{\rho^2 + a^2\cos^2\theta}$$

Since this is clearly negative, the infinite red shift surface will pass physical objects in both directions and is not a one-way membrane. We shall seek an axially symmetric and time-dependent null hypersurface

(7.132) $$u(r,\theta) = \text{const} \qquad n_\alpha = \left(0, \frac{\partial u}{\partial r}, \frac{\partial u}{\partial \theta}, 0\right)$$

A differential equation for u results from setting the norm of n_α equal to zero

(7.133) $$(\rho^2 - 2m\rho + a^2)\left(\frac{\partial u}{\partial \rho}\right)^2 + \left(\frac{\partial u}{\partial \theta}\right)^2 = 0$$

This is separable and easily solved; we set up a product solution

(7.134) $$u(\rho,\theta) = R(\rho)\Theta(\theta)$$

and find

(7.135) $$-(\rho^2 - 2m\rho + a^2)\left(\frac{\partial R/\partial\rho}{R}\right)^2 = \left(\frac{\partial\Theta/\partial\theta}{\Theta}\right)^2$$

Since the left side of this equation is a function of ρ alone and the right side a function of θ alone, both must be equal to a positive constant, which we may call λ. Thus

(7.136) $$\frac{\partial\Theta}{\partial\theta} = \sqrt{\lambda}\,\Theta \qquad \Theta = A\exp(\sqrt{\lambda}\,\theta)$$

where A is an arbitrary constant. However, this is not an acceptable solution in general since it is not periodic in θ and therefore does not correspond to a real surface unless $\lambda = 0$, which implies $\Theta = \text{const}$. With $\lambda = 0$ we then obtain

(7.137) $$\left(\frac{\partial R/\partial\rho}{R}\right)^2 (\rho^2 - 2m\rho + a^2) = 0$$

The solution $\partial R/\partial\rho = 0$ may be rejected, and we are being left with

the two solutions

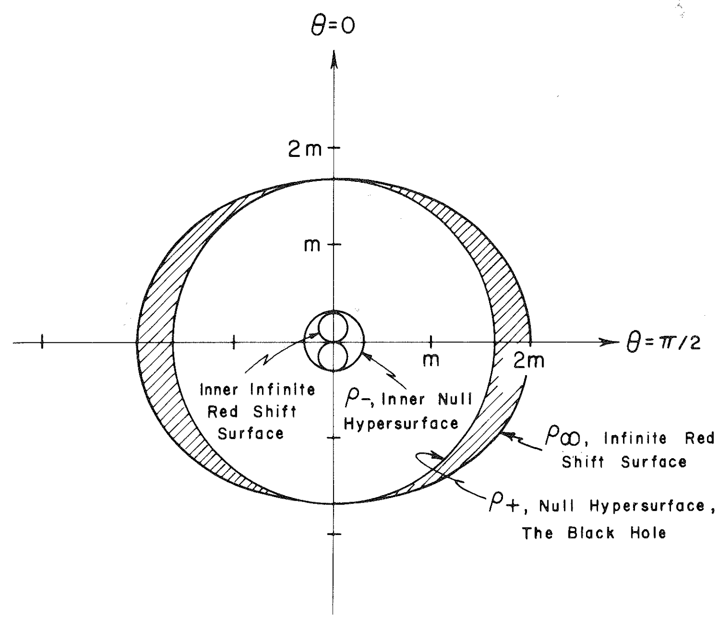
(7.138) $$\rho_\pm = m \pm \sqrt{m^2 - a^2}$$

It is remarkable that we obtain spherical surfaces and that these are well defined only so long as $|a| < m$. In the limit of $a \to 0$ these two surfaces reduce to the Schwarzschild surface $\rho = 2m$, and the origin $\rho = 0$.

The outer one-way membrane, $\rho_+ = m + \sqrt{m^2 - a^2}$, represents a dividing surface between the region of the Kerr geometry that is accessible from the distant exterior (ρ much greater than m and a) and that which is not. It is for this reason that the inner infinite red shift surface, $\rho_- = m - \sqrt{m^2 - a^2\cos^2\theta}$, is not of great physical significance; it is entirely contained within the outer one-way membrane. On the other hand, the outer infinite red shift surface is exterior to the one-way membrane, and the region between has several interesting features. Our further discussion will deal only with the *outer* one-way membrane and the outer infinite red shift surface (see Fig. 7.2).

We have shown that the infinite red shift surface does not represent a barrier to either massive test bodies or light; both may cross the surface in either direction, except at the special points $\theta = 0$ and π, where the

Fig. 7.2
The distinguished surfaces for the Kerr metric; $a = m/2$ in this illustration.



infinite red shift surface and the one-way membrane coincide. It is interesting to consider the physical significance of this for a star whose surface has approached the infinite red shift surface, $\rho = \rho_\infty$. For the special case of a spherically symmetric nonrotating star whose exterior field is described by the Schwarzschild metric we expect that atoms on the stellar surface will be at coordinate rest, that is, $dr = d\theta = d\varphi = 0$ along their world-line. For the case of a rotating star, however, this is not true; atoms on the stellar surface will in general have a velocity in the φ direction. Thus their world-lines in the Kerr metric will not correspond to coordinate rest. Since the infinite red shift surface refers explicitly to sources at rest in the Kerr metric, we may not conclude that a star whose surface approaches the infinite red shift surface will become a black hole analogous to the Schwarzschild black hole; light may escape from the surface, dependent upon its actual motion.

To further emphasize that coordinate rest is not a reasonable state for an atom in the Kerr metric let us look further into the physical interpretation of the world-lines $dr = d\theta = d\varphi = 0$. We have $ds^2 = g_{00}c^2\,dt^2$ for such a world-line. Clearly ds^2 is positive outside, zero on, and negative inside the infinite red shift surface, since it has the sign of g_{00}. All massive bodies have world-lines for which $ds^2 > 0$, however, so that coordinate rest is possible only for an atom outside the infinite red shift surface. Curiously a body at *coordinate rest* on the surface moves with *physical velocity* c since $ds^2 = 0$. (Equivalently its velocity is c in a tangent Lorentz space.)

The nature of the one-way membrane $\rho = \rho_+$ is very different from the infinite red shift surface; it represents a surface from which no light ray may emanate, regardless of the motion of the source. The radius ρ_+ may thus be regarded as the true critical size for which a rotating star becomes a black hole, analogous to the Schwarzschild black hole.

7.9 Effective Potentials and Black Hole Energetics

In classical mechanics the central-force problem can be treated as a one-dimensional problem by the introduction of an effective potential energy, which combines the usual potential energy and a "centrifugal potential energy." In this section we shall show that an analogue of the classical effective-potential-energy function can be constructed for test bodies moving in both the Schwarzschild and Kerr metrics. We shall treat the Schwarzschild metric first because of its simplicity and similarity to the classical problem; the Kerr metric is fundamentally different and possesses very interesting energetic properties which may be of great interest in astrophysical problems.

We begin with (6.82) describing a test body of rest mass μ and give a physical interpretation of the constant l. If the motion is unbounded so that the body may escape, we have asymptotically for large r

$$\dot{r}^2 = c^2l^2 - 1 \tag{7.139}$$

Moreover, we may express $\dot{r}$, using (6.81), as

$$\dot{r} = \frac{dr}{ds} = \frac{dr}{dt}\frac{dt}{ds} = \frac{dr}{dt}\,l\left(1 - \frac{2m}{r}\right)^{-1} \to vl \tag{7.140}$$

The limit $dr/dt \to v$ follows since only the r component of the velocity remains finite asymptotically, as is evident from (6.80). We may now combine (7.139) and (7.140) to obtain l in terms of the asymptotic velocity

$$cl = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} \equiv \gamma \tag{7.141}$$

where the sign of l is chosen positive for obvious convenience. This quantity γ is precisely the Lorentz contraction factor. A useful equivalent form of (7.141) follows if we observe that in the asymptotic flat space where special relativity holds the total energy E of the body is given by its rest energy μc^2 times γ. Thus we may express the constant cl as

$$cl = \frac{E}{\mu c^2} \tag{7.142}$$

where E is commonly termed the energy at infinity; it is a constant of the motion, is equal to the energy at infinity, and can therefore be defined as the general relativistic generalization of the total energy.

To define the effective potential energy function we solve (6.82) for $\mu^2\dot{r}^2$

$$\begin{aligned}\mu^2\dot{r}^2 &= \mu^2l^2c^2 - \left(1 - \frac{2m}{r}\right)\left(\frac{\mu^2h^2}{r^2} + \mu^2\right) \\ &= \frac{E^2}{c^4} - \left(1 - \frac{2m}{r}\right)\left(\frac{\mu^2h^2}{r^2} + \mu^2\right) \equiv \frac{E^2}{c^4} - \frac{V^2}{c^4}\end{aligned} \tag{7.143}$$

The function V introduced here will be referred to as the effective potential energy; the positive root is always used. We see that the difference between the total energy and the effective potential energy provides a measure of the radial velocity of the particle.

One immediate use of the effective potential energy is in relating turning points to the constants E and h. A turning point is a point where $\dot{r} = 0$ and is thus obtained as a solution of $E^2 - V^2 = 0$.

The physical interpretation of V is most clearly seen by taking the classical limit of low velocities and weak fields. We may express V up to order mass$^2/r^2$ by expanding (7.143), using the positive root:

$$V \cong \mu c^2 - \frac{\kappa M \mu}{r} + \frac{\mu^2 c^2 h^2}{2\mu r^2} \qquad \frac{\kappa M}{c^2} \equiv m \tag{7.144}$$

From (6.80) it is evident that in the classical limit $\mu c h$ is the usual angular momentum L. Thus (7.144) states that V is approximately μc^2 plus the classical effective potential energy

$$\begin{aligned} V &\cong \mu c^2 + V_{\text{cl}} \\ V_{\text{cl}} &\equiv -\frac{\kappa M \mu}{r} + \frac{L^2}{2\mu r^2} \end{aligned} \tag{7.145}$$

If we now write E as $\mu c^2 + E_{\text{cl}}$, where $E_{\text{cl}} \ll \mu c^2$ is the usual classical total energy, we can write Eq. (7.143) as

$$\begin{aligned} \mu^2 \dot{r}^2 &\cong \frac{(\mu c^2 + E_{\text{cl}})^2}{c^4} - \frac{\mu c^2 + V_{\text{cl}}}{c^4} \\ &\cong \frac{2\mu E_{\text{cl}}}{c^2} - \frac{2\mu V_{\text{cl}}}{c^2} \end{aligned} \tag{7.146}$$

or

$$E_{\text{cl}} \cong \frac{\mu (\dot{r} c)^2}{2} + V_{\text{cl}} \cong \frac{\mu}{2}\left(\frac{dr}{dt}\right)^2 + V_{\text{cl}} \tag{7.147}$$

where we have neglected small terms of second order. This is the usual classical expression involving the effective potential energy, thus justifying our statement that V defined in (7.143) is a relativistic generalization of this quantity. For cases of bounded motion we may still define E as the total energy, using the above classical reduction as motivation.

The effective-potential-energy function is useful in analyzing the motion of bodies in the Schwarzschild field as illustrated in Exercise 7.9. We shall next define an analogous function for the Kerr metric which is of fundamental importance. For simplicity we shall limit our discussion to the special case of equatorial orbits. Our discussion relies on and follows very closely that of the Schwarzschild case.

The exact equations of motion for the Kerr metric (7.110) are easily obtained, as in the Schwarzschild metric. We might expect equatorial orbits to exist due to the symmetry of the Kerr metric about the equatorial plane. Indeed if we set $\theta = \pi/2$ in the equations of motion, as we did for the Schwarzschild problem in Sec. 6.3, we obtain the following consistent set of equations for equatorial orbits:

$$\left(1 - \frac{2m}{\rho}\right) c\dot{t} - \frac{2ma}{\rho}\dot{\varphi} = cl \tag{7.148a}$$

$$\left[\rho^2 + a^2\left(1 + \frac{2m}{\rho}\right)\right]\dot{\varphi} + \frac{2ma}{\rho} c\dot{t} = h \tag{7.148b}$$

$$\begin{aligned} \left(1 - \frac{2m}{\rho}\right) c^2 \dot{t}^2 - \left(1 - \frac{2m}{\rho} + \frac{a^2}{\rho^2}\right)^{-1} \dot{\rho}^2 & \\ - \left[\rho^2 + a^2\left(1 + \frac{2m}{\rho}\right)\right]\dot{\varphi}^2 - \frac{4ma}{\rho} c\dot{t}\dot{\varphi} &= 1 \end{aligned} \tag{7.148c}$$

We can identify cl, precisely as before, as the total energy, or energy at infinity, of the test body divided by μc^2, where μ is the test-body mass; this is evident by inspection of (7.148) for very large ρ since in this limit all terms in a vanish and the preceding discussion applies unchanged.

Equations (7.148a) and (7.148b) may be easily solved for $\dot{\varphi}$ and $\dot{t}$

$$\begin{aligned} Dc\dot{t} &= \left[\rho^2 + a^2\left(1 + \frac{2m}{\rho}\right)\right] cl + \frac{2ma}{\rho} h \\ D\dot{\varphi} &= -\frac{2ma}{\rho} cl + \left(1 - \frac{2m}{\rho}\right) h \\ D &\equiv a^2 + \rho^2\left(1 - \frac{2m}{\rho}\right) \end{aligned} \tag{7.149}$$

The above expressions may be substituted into (7.148c) and an equation for $\mu^2 \dot{\rho}^2$ obtained; after elementary manipulation the result is

$$\begin{aligned} \mu^2 \dot{\rho}^2 = \frac{1}{\rho^3}\Bigg\{ & \frac{[\rho^3 + a^2(\rho + 2m)]E^2}{c^4} + \frac{(2m - \rho)L^2}{c^2} \\ & + \frac{4maLE}{c^3} - \mu^2\rho^2(\rho - 2m) - a^2\rho\mu^2 \Bigg\} \end{aligned} \tag{7.150}$$

where, as before, we define $L = \mu h c$. This equation is the generalization of (7.143), which is used to define the Schwarzschild effective potential;

for $a = 0$ they are identical. In the present case a cross term occurs, and we may define two effective potentials by writing (7.150) as

$$\mu^2\dot{\rho}^2 = \frac{1}{c^4}(E - V_+)(E - V_-) \tag{7.151}$$

where $V_\pm$ are the roots of the polynomial in E that constitutes the right side of (7.150). These are explicitly

$$V_\pm = \frac{-2maLc \pm [\rho^2 + a^2 - 2m\rho]^{1/2}[c^2\rho^2L^2 + \mu^2\rho c^4(\rho^3 + a^2\rho + 2ma)]^{1/2}}{\rho^3 + a^2(\rho + 2m)} \tag{7.152}$$

For large ρ, $V_\pm \to \pm\mu$; thus to generalize the classical results and the Schwarzschild effective potential we choose the positive sign and identify V_+ as the physically significant effective potential. In the limit of small a we readily obtain

$$V_+ \cong V_s + \frac{2maLc}{\rho^3} \tag{7.153}$$

where V_s is the Schwarzschild effective potential. There is no direct classical analogue of the extra term on the right side of (7.153).

The Schwarzschild effective potential is positive definite for $r > 2m$. It is extraordinary that the Kerr effective potential does not possess this feature. Indeed, at the black-hole radius, where $\rho^2 + a^2 - 2m\rho = 0$, V_+ is given by

$$V_+ = \frac{-2maLc}{\rho^3 + a^2(\rho + 2m)} \tag{7.154}$$

which will be negative if a and L have the same sign. Since a has the opposite sign of the black-hole angular momentum J by (7.115), this means that the test body has the opposite sign of angular momentum L, that is, is counterrotating. This may be interpreted physically as follows: a body released with $\dot{\rho} = 0$ at ρ will have total energy $E = V_+(\rho)$. This total energy can be negative for the Kerr metric, which corresponds to a binding energy in excess of the rest energy and leads to the possibility of interesting physical effects, as we shall presently discuss. Let us return to (7.152) and ask for what values of ρ the effective potential V_+ can be negative. Clearly the mass term μ^2 contributes positively to V_+, so to seek minimum V_+ we consider the limit of $\mu \to 0$. Then V_+ will be zero or negative when

$$(\rho^2 - 2m\rho + a^2)\rho^2L^2 \le 4m^2a^2L^2 \tag{7.155}$$

The maximum solution for zero energy is $\rho = 2m$. We therefore see that a test body of zero mass can have a negative total energy for ρ between $m + \sqrt{m^2 - a^2}$ and $2m$.

The above analysis may be generalized to nonequatorial orbits, in which case one would find that negative total energy is possible between the null or black-hole surface $\rho = m + \sqrt{m^2 - a^2}$ and the infinite red shift surface $\rho = m + \sqrt{m^2 - a^2\cos^2\theta}$. This region is commonly called the *dynamic zone* or *ergosphere*. The infinite red shift surface, which previously had little physical significance, is now seen to play an important role in the energetics of the black hole.

By using orbits entering the ergosphere, energy can be extracted from a Kerr black hole (Penrose, 1969). One sends a particle into the ergosphere with energy E_1 and lets it decay into two particles with energies E and E_2 in such a way that E is negative, as we have shown is possible. Then overall conservation of energy implies that E_2 will be larger than E_1. The rotational energy of the black hole must in general decrease in the process. This raises the possibility that rotating black holes may be sources of large amounts of energy. If the energy can be extracted in the form of electromagnetic or gravitational radiation, this would be of great interest in the study of quasars, whose energy source remains a mystery.

Exercises

7.1 Explicitly verify the general properties of a degenerate metric (7.8), (7.9), and $g = -1$ for Eddington's special case (7.4).

7.2 Show that the order m^4 equations (7.15d) are satisfied identically, and verify also Eq. (7.16).

7.3 In the text $l_0^2 = \alpha$ is shown to be a solution of (7.25). To show that it is unique write the general solution as $l_0^2 = f\alpha$, with f an arbitrary function, and show by steps analogous to those following (7.65) that $\nabla f = 0$. Thus f must be a constant, and l_0^2 is unique up to this multiplicative constant.

7.4 Prove that

$$A = \frac{\beta^2 - \alpha^2}{2l_0} \qquad \text{and} \qquad L = \frac{\beta^2 + 3\alpha^2}{2l_0}$$

7.5 Show that no loss of generality occurs if the displacement discussed in Sec. 7.5 is assumed to have the form (7.86); do this by beginning with general complex a_i in (7.85) and then performing real physical rotations and translations to obtain (7.86).

7.6 From (7.98) and (7.99) compute $\cos^2\theta + \sin^2\theta$ and use (7.89) to show that it is 1, thereby showing that θ is a reasonable angular coordinate.

7.7 The dimensionless quantities $\alpha_1 = m/r$ and $\alpha_2 = \kappa J/c^3r^2$ occur in the Lense-Thirring metric as expansion coefficients. Show that for a reasonable star the ratio $\alpha_2/\alpha_1 < 1$ outside the surface. Show also that for a star of reasonable mass and rapid rotation one may have $\alpha_1^2/\alpha_2 \ll 1$. Thus there exist situations where the first-order approximations of the Lense-Thirring metric are justified.

7.8 Discuss the positions of the singularities of γ for the Kerr and Schwarzschild metrics. How do the singularities relate to the symmetry of the fields?

7.9 Use the Schwarzschild potential-energy function to show that the minimum radius for stable circular orbits is $r = 6m$.

7.10 In general the Kerr metric allows no radial null geodesics unless $a = 0$; show this, and consider the exceptional case of a null geodesic along the z axis of axial symmetry. Then consider null geodesics which are radial for large r and study their behavior as a function of r, that is, how do they deviate in the φ direction for small r? What can be said about the light-trapping properties of a rotating black hole versus a nonrotating black hole?

7.11 Consider a metric of the form

$$ds^2 = f(r)c^2\,dt^2 - \frac{dr^2}{f(r)} - r^2(d\theta^2 + \sin^2\theta\,d\varphi^2)$$

and show that it can be put into the degenerate form considered in this chapter by a coordinate transformation involving only the time.

Problems

7.1 Show, using the degenerate metric form (7.5), that

$$R^\alpha{}_{\beta\gamma\delta}l^\beta l^\delta = -2mA^2 l_\gamma l^\alpha$$

and thus that

$$R_{\alpha\beta\delta\gamma}l_\omega l^\beta l^\delta - R_{\alpha\beta\delta\omega}l_\gamma l^\beta l^\delta = 0$$

Penrose (1960) has shown that this equation implies that the space-time is algebraically special, i.e., it is not of Petrov type I.

7.2 In the study of geometrical optics in a Riemann space, quantities called the *expansion*, the *twist*, and the *shear* of a family of light rays characterize the shadow cast by a two-dimensional disk on a nearby screen. Study these quantities (Pirani, 1965) and show that for the present situation the expansion is $-\alpha$, the twist is β, and the shear is zero.

7.3 As noted in the text, coordinate rest is not a possible condition for a physical object inside the infinite red shift surface. Show that at any given point there exists a coordinate transformation from t and φ to $\hat{t}$ and $\hat{\varphi}$ such that coordinate rest in terms of $\hat{t}$ is possible for a physical object. Thus show that t is not a good time marker inside the infinite red shift surface but $\hat{t}$ is (see Vishveshwara, 1968).

7.4 Show that there exist physically acceptable trajectories which cross the infinite red shift and continue to spatial infinity (see Carter, 1968).

7.5 The Penrose process discussed in Sec. 7.9 is potentially capable of extracting large amounts of energy from a rotating black hole. Study this process further and discuss the mechanism and efficiency of extraction (Penrose, 1969; Christodolou, 1970). In particular show that energy of the order of one-third the total mass energy may be extracted from a rapidly spinning black hole.

7.6 We studied equatorial orbits in the text to understand the energy properties of the dynamic zone, or ergosphere. Study the work of Carter (1968), in particular the implications of the so-called Killing tensor, and generalize the discussion to show that there exist negative energy orbits in the entire ergosphere.

7.7 The metric form considered in this chapter corresponds to an algebraically special space-time, as noted in Exercise 7.11 and Prob. 7.1. Show that not all algebraically special space-times have a metric of this form.

7.8 What is the Petrov type of the Kerr solution? (See also Prob. 7.1.)

Bibliography

Boyer, R. H., and R. W. Lindquist (1967): Maximal Analytic Extension of the Kerr Metric, *J. Math. Phys.*, **8**:265.

Carter, B. (1968): Global Structure of the Kerr Family of Gravitational Fields, *Phys. Rev.*, **174**:1559.

Christodolou, D. (1970): Reversible and Irreversible Transformations in Black-Hole Physics, *Phys. Rev. Letters*, **25**:1596.

Cohen, J. M. (1968): Angular Momentum and the Kerr Metric, *J. Math. Phys.* **9**:905.
Debney, G. C., R. P. Kerr, and A. Schild (1969): Solutions of the Einstein-Maxwell Equations, *J. Math. Phys.*, **10**:1842.
Eddington, A. S. (1924): A Comparison of Whitehead's and Einstein's Formulae, *Nature*, **113**:192.
Hawking, S. W., and R. Penrose (1970): The Singularities of Gravitational Collapse and Cosmology, *Proc. Roy. Soc. London*, **314**:529.
Kerr, R. P. (1963): Gravitational Field of a Spinning Body as an Example of Algebraically Special Metrics, *Phys. Rev. Letters*, **11**:237.
Kerr, R. P., and A. Schild (1965): Some Algebraically Degenerate Solutions of Einstein's Gravitational Field Equations, in "Applications of Nonlinear Partial Differential Equations in Mathematical Physics," *Proc. Symp. Appl. Math., American Math. Soc.*, **18**:199.
Lense, J., and H. Thirring (1918): Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie, *Phys. Z.*, **19**:156.
Penrose, R. (1960): A Spinor Approach to General Relativity, *Ann. Phys.*, **10**:171.
Penrose, R. (1969): Gravitational Collapse: The Role of General Relativity, *Riv. Nuovo Cimento*, **1**:252.
Pirani, F. A. E. (1965): "Introduction to Gravitational Radiation Theory," Lectures in General Relativity, Brandeis Summer Institute in Theoretical Physics, vol. I, Englewood Cliffs, N.J., 1964.
Ruffini, R., and J. A. Wheeler (1971): Relativistic Cosmology and Space Platforms, *Proc. Conf. Space Phys., ESRO Paris Meeting.*
Schiffer, M. M., R. J. Adler, J. Mark, and C. Sheffield (1973): Kerr Geometry as Complexified Schwarzschild Geometry, *J. Math. Phys.*, **14**:52.
Vishveshwara, C. V. (1968): Generalization of the "Schwarzschild Surface" to Arbitrary Static and Stationary Metrics, *J. Math. Phys.*, **9**:1319.

CHAPTER

8

The Mathematical Structure of the Einstein Differential System; the Problem of Cauchy

Einstein's equations in a matter-free region form a system of 10 second-order quasi-linear differential equations in the four space-time variables. These equations are to be solved for the 10 unknown components of the metric tensor $g_{\alpha\beta}$, which we may interpret as gravitational potentials. The equations are

$$R_{\mu\nu} = 0 \tag{8.1}$$

or equivalently, by the definition (5.119),

$$\begin{Bmatrix} \sigma \\ \mu\ \sigma \end{Bmatrix}_{|\nu} - \begin{Bmatrix} \sigma \\ \mu\ \nu \end{Bmatrix}_{|\sigma} + \begin{Bmatrix} \alpha \\ \mu\ \sigma \end{Bmatrix}\begin{Bmatrix} \sigma \\ \alpha\ \nu \end{Bmatrix} - \begin{Bmatrix} \alpha \\ \mu\ \nu \end{Bmatrix}\begin{Bmatrix} \sigma \\ \alpha\ \sigma \end{Bmatrix} = 0 \tag{8.2}$$

In this form the quasi-linear nature of the equations is apparent; only the two first terms contain the second derivatives of the $g_{\alpha\beta}$'s, and they are linear in them, whereas they and the other terms are obviously not linear in the first derivatives or in the $g_{\alpha\beta}$'s themselves.

Equations (8.2) connect the time derivatives of the components of the metric tensor $g_{\mu\nu}$ with the space derivatives and the $g_{\mu\nu}$ themselves. They allow the solution of the following mathematical problem: Given the metric tensor and all its first derivatives at a given moment x^0 in the entire three-dimensional space of the remaining three variables x^i, to compute its value for all future time. This is a typical initial-value problem in the theory of partial differential equations and represents the causal development of a physical system from initial data. This fundamental problem in the mathematical theory of partial differential equations is known as the "Cauchy problem" after one of its first investigators. It stands in complete analogy to the general problem of a classical mechanical system whose evolution is determined by the initial positions and velocities of its elements. We wish to give in this chapter a detailed study of this initial-value problem, which will throw considerable light on the qualitative structure of the field equations.

8.1 Formulation of the Initial-Value Problem

Let us prescribe a three-dimensional hypersurface S oriented in space. We can then choose, without any loss of generality, a coordinate system such that the hypersurface S is described by the equation $x^0 = 0$. Since the normal to a surface oriented in space is itself oriented in time, our assumption forces g_{00} to be positive. Physically, we may interpret S as representing the space at the given time $x^0 = 0$. In this space we prescribe the components of the metric tensor $g_{\mu\nu}$ and their first derivatives. However, giving the values of the metric potentials all over S automatically allows us to compute all their derivatives taken in S (interior derivatives). That is, the derivatives $g_{\mu\nu|i}$, which do not contain any differentiation with respect to the index 0 (time), are known on S by differentiation of the given $g_{\mu\nu}$ in S. Therefore it is sufficient to give only the following initial values on S:

(8.3) $g_{\mu\nu}\,,\ g_{\mu\nu|0}$ (metric potentials and their normal derivatives)

These initial data, together with the differential system (8.1), form a typical Cauchy initial-value problem in the theory of partial differential equations, which we shall now discuss in detail.

8.2 Structure of Einstein's Equations

To investigate the Cauchy problem for Einstein's equations, we proceed in a standard manner always followed in initial-value problems for partial differential equations. We try to express the second time derivatives of our unknown functions in terms of the known space derivatives and the first-order time derivatives. If this can be done in analytic form, we may differentiate these identities indefinitely in time and obtain recursively all time and space derivatives of the $g_{\mu\nu}$ in S in terms of the given initial data. If the $g_{\mu\nu}$ admit a power-series development in x^0 near to S, we shall be able to compute all coefficients of this development by the above procedure. Thus this development will be uniquely determined by our initial data. Once one has achieved a continuation of the solution from an initial surface S into the future, one can try to study the evolution by indefinite continuation for all time and space.

Let us write out the contracted Riemann tensor in a form which displays clearly all the second derivatives of the metric tensor:

$$R_{\mu\nu} = \tfrac{1}{2}g^{\rho\sigma}[-g_{\mu\sigma|\nu|\rho} - g_{\nu\rho|\mu|\sigma} + g_{\mu\nu|\rho|\sigma} + g_{\rho\sigma|\mu|\nu}] + K_{\mu\nu} \tag{8.4}$$

Here the term $K_{\mu\nu}$ contains only the metric potentials and their first derivatives. It is due to the last two terms in (8.2) and to the remainder of terms in the first two Christoffel symbols which do not contain derivatives higher than first order. Hence $K_{\mu\nu}$ is known on S from the initial data. By differentiation in S, that is, in the space variables x^i, all higher derivatives of the $g_{\mu\nu}$ can be computed on S, as long as they do not involve more than one differentiation in time. All such derivatives are expressible by successive differentiation of the initial data on S, which must of course be assumed to be indefinitely differentiable.

We can next use Eqs. (8.1) and the form (8.4) of $R_{\mu\nu}$ to compute the second time derivatives of the metric potentials in terms of the known data on S. We observe that the indices $i, j = 1, 2, 3$ appear quite differently from the time index 0. An easy calculation shows

$$R_{ij} = \tfrac{1}{2}g^{00}g_{ij|0|0} + M_{ij} = 0 \tag{8.5a}$$

$$R_{i0} = -\tfrac{1}{2}g^{0j}g_{ij|0|0} + M_{i0} = 0 \tag{8.5b}$$

$$R_{00} = \tfrac{1}{2}g^{ij}g_{ij|0|0} + M_{00} = 0 \tag{8.5c}$$

where the $M_{\mu\nu}$ can be expressed in terms of the initial data on S.

We find the following surprising situation:

1. The linear system (8.5) for the calculation of the second time derivatives does not contain the unknowns $g_{\lambda 0|0|0}$ which are needed to determine the time evolution of the metric from the data on S. We have a problem of *underdetermination*.
2. The linear system (8.5) represents a set of 10 equations for the six unknowns $g_{ij|0|0}$ on S, which presents a problem of *overdetermination* and leads to compatibility requirements for the data $M_{\mu\nu}$ on S.

Let us start with problem 1. It is not surprising that the knowledge of $R_{\mu\nu} = 0$ does not determine the $g_{\mu\nu}$ in a unique way. Even in a flat space for which the entire Riemann tensor vanishes, the metric potentials $g_{\mu\nu}$ are still somewhat arbitrary, depending on our choice of coordinates. We shall now show that we can always make a coordinate transformation in the neighborhood of S such that the $g_{\mu\nu}$ and their first derivatives in S are unchanged but $g_{\lambda 0|0|0} \equiv 0$ on S. This fact explains why Eqs. (8.1) could not possibly contain information on these second time derivatives; indeed, the change of coordinates does not affect the tensor relations (8.1), which must be valid in all equivalent coordinate systems.

We introduce the coordinate transformation

$$\bar{x}^\lambda = x^\lambda + \tfrac{1}{6}(x^0)^3 A^\lambda(x^\alpha) \tag{8.6a}$$

where the $A^\lambda(x)$ represent four functions defined in the neighborhood of S which will be chosen to fit our requirements. Under the transformation (8.6a), the equation of S remains obviously $\bar{x}^0 = 0$, and we have on S

$$\frac{\partial \bar{x}^\lambda}{\partial x^\mu} = \delta^\lambda{}_\mu \qquad \left(\frac{\partial \bar{x}^\lambda}{\partial x^\mu}\right)_{|\nu} = 0 \qquad \text{on } S \tag{8.6b}$$

Even all second derivatives of the Jacobi matrix $\partial\bar{x}^\lambda/\partial x^\mu$ vanish on S except for

$$\left(\frac{\partial \bar{x}^\lambda}{\partial x^0}\right)_{|0|0} = A^\lambda(x^\alpha) \tag{8.6c}$$

We observe that the coordinate transformation (8.6a) has a nonvanishing Jacobian near S and is therefore an admissible transformation.

We now apply (8.6b) and (8.6c) to find the relation between the metric potentials and their first two derivatives in the two coordinate systems on the surface S. We have the identity

$$g_{\lambda\mu} = \bar{g}_{\alpha\beta}\frac{\partial \bar{x}^\alpha}{\partial x^\lambda}\frac{\partial \bar{x}^\beta}{\partial x^\mu} \tag{8.7}$$

and since the Jacobi matrix behaves on S like the unit matrix up to the second derivatives, we find *on* S

$$g_{\lambda\mu} = \bar{g}_{\lambda\mu} \qquad g_{\lambda\mu|\nu} = \bar{g}_{\lambda\mu|\nu} \qquad g_{\lambda\mu|\nu|\rho} = \bar{g}_{\lambda\mu|\nu|\rho} \tag{8.8}$$

except for the case $\nu = \rho = 0$. In this case, (8.6b) and (8.7) yield on S

$$g_{\lambda\mu|0|0} = \bar{g}_{\lambda\mu|0|0} + \bar{g}_{\alpha\mu}\left(\frac{\partial \bar{x}^\alpha}{\partial x^\lambda}\right)_{|0|0} + \bar{g}_{\lambda\beta}\left(\frac{\partial \bar{x}^\beta}{\partial x^\mu}\right)_{|0|0} \tag{8.9}$$

The last two terms in (8.9) vanish except for either λ or μ being zero. Using (8.6c), we thus arrive at the equations

$$\begin{aligned} g_{ij|0|0} &= \bar{g}_{ij|0|0} \qquad g_{i0|0|0} = \bar{g}_{i0|0|0} + \bar{g}_{i\beta}A^\beta \\ g_{00|0|0} &= \bar{g}_{00|0|0} + 2\bar{g}_{0\beta}A^\beta \end{aligned} \tag{8.10}$$

We choose now the as yet undetermined functions $A^\beta(x)$ in such a way that all $\bar{g}_{\lambda 0|0|0}$ vanish on S. For this purpose, we have to demand only that

$$g_{i0|0|0} = g_{i\beta}A^\beta \qquad \tfrac{1}{2}g_{00|0|0} = g_{0\beta}A^\beta \tag{8.11}$$

since we know that on S we have $g_{\alpha\beta} = \bar{g}_{\alpha\beta}$. The system (8.11) consists of four linear equations for the four unknowns A^β with the determinant $g = \det((g_{\mu\nu}))$ of the metric tensor itself. This determinant is nonzero, since we always assume that the metric is everywhere regular. Thus the $A^\beta(x)$ are determined by (8.11) at every point on S in a unique way.

Our purely formal analysis of the equation system (8.1) has led to an important insight into the causal meaning of these equations. The initial data on S do not determine the resulting metric in a unique way; the solution contains the four arbitrary functions $g_{\lambda 0|0|0}$ which are at our disposal. It should be observed that this arbitrariness is due to the fact that we can pick an arbitrary coordinate system for the description of the space-time continuum. However, the solutions obtained will differ only formally; they will describe the same geometrico-physical situation in different reference systems. This feature of the Einstein equations (8.1) was already stressed by Hilbert in 1915. He drew an important conclusion from this fact (Hilbert, 1915): Since the 10 differential equations (8.1) leave a freedom of four arbitrary functions in the solution, they cannot be entirely independent, but must have four inner relations. These relations are, of course, a consequence of the Bianchi identities for the full Riemann tensor and are known to us from Chap. 5 in the form of the condition that the Einstein tensor be divergenceless.

In the following we shall suppose that on the initial-value surface S we have chosen the metric tensor in such a way that $g_{\lambda 0|0|0} \equiv 0$. This normalization removes the first difficulty mentioned above of underdetermination in the field equations (8.1).

We return now to the system of differential equations (8.5), which contains 10 conditions for the six unknowns $g_{ij|0|0}$. We face here the problem of overdetermination. We observe that the first set of equations $R_{ij} = 0$ is sufficient to calculate the $g_{ij|0|0}$ and that the additional four equations $R_{i0} = 0$ and $R_{00} = 0$ must therefore be necessary consequences of this determination. To study this question in detail, we combine the first set of equations (8.5) with the second and third in the form

$$g^{00}R_{i0} + g^{0j}R_{ij} = g^{00}M_{i0} + g^{0j}M_{ij} = 0$$
$$g^{00}R_{00} - g^{ij}R_{ij} = g^{00}M_{00} - g^{ij}M_{ij} = 0 \tag{8.12}$$

We see that these combinations, which must be zero by virtue of the field equations (8.1), depend only on the values of the components of the metric tensor and their first derivatives and upon the derivatives of these quantities with respect to the space variables. Hence Eqs. (8.12) represent four constraints on the initial data on S. As usual, if a system of equations does not determine a unique solution, one runs into a set of compatibility conditions for the data in order that a solution may be possible at all. Equations (8.12) are of this nature.

The structure of Eqs. (8.5) and (8.12) becomes very clear if we consider, besides the contracted Riemann tensor $R_{\alpha\beta}$, the Einstein tensor $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta}$. We have, by definition and (8.12),

$$G_i{}^0 = R_i{}^0 = g^{0j}R_{ij} + g^{00}R_{i0} = g^{00}M_{i0} + g^{0j}M_{ij} \tag{8.13}$$

$$G_0{}^0 = R_0{}^0 - \tfrac{1}{2}R = g^{0\alpha}R_{0\alpha} - \tfrac{1}{2}(g^{ij}R_{ij} + g^{0i}R_{0i} + g^{i0}R_{i0} + g^{00}R_{00}) \tag{8.14}$$

which simplifies to

$$G_0{}^0 = \tfrac{1}{2}(g^{00}R_{00} - g^{ij}R_{ij}) = \tfrac{1}{2}(g^{00}M_{00} - g^{ij}M_{ij}) \tag{8.14$'$}$$

Thus the equation system (8.5) may be written in the equivalent form

$$R_{ij} = 0 \tag{8.15a}$$

$$G_\lambda{}^0 = G^0{}_\lambda = 0 \tag{8.15b}$$

This normal form of the field equations is due to Lichnerowicz (1955). The set (8.15a) of six equations serves to determine the six unknown functions $g_{ij|0|0}$ from the initial data on S. The additional four equations (8.15b) in terms of the Einstein tensor $G_\lambda{}^0$ depend only on the initial data and represent necessary conditions on the initial data in order that a solution exist at all.

We have shown earlier that the differential equations for the metric field in empty space may be written in the two equivalent tensor forms

$R_{\mu\nu} = 0$ or $G_{\mu\nu} = 0$. The new system (8.15) is not covariant and is formally inferior to the other two formulations. However, in a fixed coordinate system it is particularly convenient for the study of the initial-value problem, as we shall see in the next section.

8.3 Separation of the Cauchy Problem into Two Parts

So far, we have given simple formulas for all second derivatives of the metric tensor on S as follows:

$g_{ij\|0\|0} = \dfrac{-2M_{ij}}{g^{00}}$	from (8.5) since $g^{00} > 0$
$g_{\lambda 0\|0\|0} = 0$	by normalization of solution
$g_{\alpha\beta\|i\|\lambda}$	known by interior differentiation of initial data on S

On the other hand, we found that the initial data on S must satisfy the compatibility condition $G^0{}_\alpha = 0$.

If we wish to extend the solution into space-time outside of the hypersurface S, we have to use the equation system (8.5), or equivalently the two sets of equations (8.15a) and (8.15b). We shall now prove a remarkable fact, namely, that once equations (8.15b) are satisfied on the initial hypersurface S, they will remain automatically valid for all time. Thus Eqs. (8.15b) are essentially a consequence of the more fundamental system (8.15a). We may interpret $G^0{}_\alpha = 0$ as integrals of the differential system (8.15a).

To prove our assertion, we start with the Einstein tensor, which is by definition $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta}$, and use the fact that, by virtue of (8.15a), all R_{ij} vanish. Hence we easily find

$$G^i{}_j = g^{i0}R_{0j} - \tfrac{1}{2}g^i{}_j(g^{00}R_{00} + 2g^{0l}R_{0l})$$
$$G^i{}_0 = g^{i0}R_{00} + g^{il}R_{l0} \tag{8.16}$$

and

$$G^0{}_j = g^{00}R_{0j} \qquad G^0{}_0 = \tfrac{1}{2}g^{00}R_{00} \tag{8.17}$$

Observe that the right-hand sides in (8.16) depend only on $R_{0\lambda}$ and can be expressed by means of (8.17) in terms of $G^0{}_\lambda$. We see that $G^i{}_\lambda$ depends linearly upon $G^0{}_\lambda$, with coefficients which are easily expressed in terms of the metric tensor.

We next use the fact that the Einstein tensor has a zero divergence as was shown in Chap. 5. Thus we have

$$G^{\alpha}{}_{\lambda\|\alpha} = G^{0}{}_{\lambda\|0} + G^{i}{}_{\lambda\|i} = 0 \tag{8.18}$$

If we now express the $G^i{}_\lambda$ in terms of the $G^0{}_\lambda$, we obtain in (8.18) a differential system for the $G^0{}_\lambda$ only. Indeed, carrying out the covariant differentiations in (8.18) and rearranging terms, we arrive at the form

$$G^{0}{}_{\lambda|0} = A^{\sigma i}{}_{\lambda}G^{0}{}_{\sigma|i} + B^{\sigma}{}_{\lambda}G^{0}{}_{\sigma} \tag{8.19}$$

where the coefficients $A^{\sigma i}{}_\lambda$ and $B^\sigma{}_\lambda$ depend only on the metric tensor and its first derivatives.

We have in (8.19) a system of four linear homogeneous partial differential equations of the first order for the four components $G^0{}_\lambda$. This system is already in normal form with respect to the variable x^0; indeed, the time derivatives of the unknown functions stand on the left, the functions and their spatial derivatives on the right. The coefficients of this system are by hypothesis continuous in the region considered. Hence the initial-value problem for this system possesses a unique solution for given data on S. In particular, we see that the initial data $G^0{}_\lambda = 0$ on S imply $G^0{}_\lambda \equiv 0$ since this is clearly a solution of our initial-value problem. Thus the general theory of the Cauchy problem for linear partial differential equations has led to the asserted dependence of the system (8.15*b*) upon the system (8.15*a*).

We have now obtained a clearer insight into the above-mentioned assertion of Hilbert about the four relations which must prevail between the 10 equations of the Einstein system (8.1). These relations are most clearly seen in the dependence of (8.15*b*) on (8.15*a*) and are an obvious consequence of the Bianchi identities and the related properties of the Einstein tensor.

Now the computational aspect of the system (8.15*a*) and (8.15*b*) should be quite obvious. We may prescribe on S initial data $g_{\alpha\beta}$ and $g_{\alpha\beta|\lambda}$ which are compatible with the condition $G^0{}_\lambda = 0$ on S. Next, we prescribe the four components $g_{\alpha 0}$ quite arbitrarily in time and space, subject only to the condition that they match the data on S. (See Sec. 2.4 on Gaussian coordinates.) For the sake of normalization, we demand also $g_{\alpha 0|0|0} = 0$ on S. With the $g_{\alpha 0}(x)$ so chosen, we return to the system

$$R_{ij} = \tfrac{1}{2}g^{00}g_{ij|0|0} + M_{ij} = 0 \tag{8.20}$$

where the M_{ij} depend on the given $g_{\alpha 0}(x)$ and their derivatives and on the unknowns g_{ij} and their derivatives. However, M_{ij} does not contain

time derivatives of order greater than 1. We now have a proper initial-value problem for the $g_{ij}(x)$, which can be handled by the standard methods of the theory. The g_{ij} and $g_{\alpha 0}$ then lead to an Einstein tensor $G_{\alpha\beta}$, which satisfies $G^0{}_\lambda = 0$, as we have shown above.

Since $R_{ij} = 0$ is likewise fulfilled, we have shown that we can solve the initial-value problem locally while prescribing the unessential components $g_{\alpha 0}(x)$ quite arbitrarily.

The important insight gained by our analysis is that we cannot prescribe freely the metric on a spacelike hypersurface and obtain an evolution in time of these initial data by means of the Einstein equations. On the contrary, we have the inner compatibility conditions $G^0{}_\lambda = 0$ in the initial surface $x^0 = 0$. The Einstein equations impose these conditions on a three-space which is supposed to be empty of matter. On the other hand, once these conditions are fulfilled, an evolution can take place which can be calculated from the six Einstein equations $R_{ik} = 0$. These equations determine the development of the geometry in time within the arbitrary assignment of the $g_{0\lambda}$. They also guarantee that the compatibility conditions $G^0{}_\lambda = 0$ remain true once they are fulfilled for the time $x^0 = 0$. We have thus split the Einstein equation system into two parts with essentially different significance: (1) The condition that space be empty in the spacelike hypersurface $x^0 = \text{const}$ demands $G^0{}_\lambda = 0$. (2) The equations of time evolution of the geometry $R_{ik} = 0$ determine the future development of each compatible metric in the initial space.

The question of actually determining a solution g_{ij} with given initial data and chosen $g_{\alpha 0}$ belongs to the theory of the Cauchy problem and has great difficulties of its own. We have carried out the reduction of the rather involved field equations for the metric field in empty space to a situation where the purely mathematical investigation can proceed in standard fashion.

One can consider the Cauchy problem for the Einstein system to consist of two parts: We must first verify that the compatibility conditions $G^0{}_\lambda = 0$ are satisfied by the initial data on S. Then, in principle, we have only to solve the six second-order equations $R_{ij} = 0$ throughout the four-dimensional space since the first-order system $G^0{}_\lambda = 0$ will then be identically satisfied. But when one attempts to define a procedure to construct a solution as we did above, one has to rely on both the second-order system $R_{ij} = 0$ and the first-order system $G^0{}_\lambda = 0$ simultaneously. The proof of the existence and of the uniqueness of the solutions of Eqs. (8.15*a*) has been given under a certain simple differentiability hypothesis by Fourès-Bruhat (1952). A much easier proof can be given if one assumes analyticity of the solutions and uses the Cauchy-Kowalewsky theorem. But the assumption of analyticity is unnecessary and unnatural; indeed, we shall see in the next section that Einstein's equations are of hyperbolic type and thus do not require analytic solutions. A very clear discussion of this last point in the general case of second-order partial differential equations is given by Hadamard (1932).

8.4 Characteristic Hypersurfaces of the Einstein Equation System

As usual, the best insight into the nature of a system of differential equations will be obtained by studying those singular hypersurfaces for which the Cauchy initial-value problem cannot be solved without restriction. These are the characteristic hypersurfaces of the system. As was already pointed out in Chap. 4, along such hypersurfaces different solutions of the same equation system can meet continuously, and for this reason the characteristic hypersurfaces play the role of wave fronts in the propagation of physical phenomena and are the locus in space-time of signals carried by this phenomenon.

To find these singular hypersurfaces S, we must ask that at each point of S the determination of $g_{ij|0|0}$ in terms of the Cauchy data on S be impossible. This is clearly the case if and only if at each point of S we have $g^{00} = 0$. However, this description is not covariant and is valid only in a coordinate system such that the equation for S is $x^0 = \text{const}$. If we use another coordinate system $\bar{x}^\alpha$, we have, by the transformation law of tensors,

$$\bar{g}^{\alpha\beta} \frac{\partial x^0}{\partial \bar{x}^\alpha} \frac{\partial x^0}{\partial \bar{x}^\beta} = g^{00} = 0 \tag{8.21}$$

This is a partial differential equation for the characteristic hypersurface S which has the equation

$$x^0 = \varphi(\bar{x}^\alpha) = \text{const} \tag{8.22}$$

What is the geometric meaning of this characteristic equation? We observe that

$$\bar{\xi}_\alpha = \varphi_{|\alpha} = \frac{\partial x^0}{\partial \bar{x}^\alpha} \tag{8.23}$$

is a covariant vector. By virtue of (8.21), we have

$$\bar{g}^{\alpha\beta} \bar{\xi}_\alpha \bar{\xi}_\beta = 0 \tag{8.24}$$

that is, $\bar{\xi}_\alpha$ is a null vector. If $d\bar{x}^\alpha$ is a tangent vector to the characteristic surface S with the equation $\varphi = 0$, we have on S

$$d\varphi = \varphi_{|\alpha}\, d\bar{x}^\alpha = \bar{\xi}_\alpha\, d\bar{x}^\alpha = 0 \tag{8.25}$$

that is, $\bar{\xi}_\alpha$ is the normal vector to S. Thus the characteristic hypersurfaces S of the Einstein field equations are characterized by the fact that their normal vector is at every point a null vector. (We have already studied such null hypersurfaces in connection with their "one-way membrane" properties in Chap. 7.)

Consider next the vector $\bar{\xi}^\alpha$, the contravariant form of the normal vector $\bar{\xi}_\alpha$. A differential $d\bar{x}^\alpha$ in the direction $\bar{\xi}^\alpha$ is orthogonal to the normal vector by virtue of (8.24) and lies, therefore, in S. Thus we can also infer that S possesses a tangent null vector at every point. Through every point $\bar{x}$ of a hypersurface S we can draw the cone of vectors $d\bar{x}^\alpha$ which satisfy the equation $g_{\alpha\beta}\, d\bar{x}^\alpha\, d\bar{x}^\beta = 0$, the local light cone at $\bar{x}$. We see from the above that a characteristic hypersurface touches a local light cone at all its points since the differential $d\bar{x}^\alpha$ in the direction $\bar{\xi}^\alpha$ lies in S and also lies on the local light cone by virtue of (8.24).

8.5 Bicharacteristics of the Einstein System

We found in the preceding section that the characteristic surfaces S of the Einstein equations can be described in the form $\varphi(x^\alpha) = \text{const}$, where the function $\varphi(x^\alpha)$ satisfies the first-order partial differential equation

$$H(x^\mu, \varphi_{|\nu}) = g^{\alpha\beta}\varphi_{|\alpha}\varphi_{|\beta} = 0 \tag{8.26}$$

The general theory of surfaces which satisfy first-order partial differential equations asserts that such surfaces can be built up from a family of elements, the so-called strips, which depend only on one parameter. In order to give a precise definition of a strip, we first need the concept of a surface element. A surface element of S is characterized by its location in space-time x^α and by its normal vector $p_\alpha = \varphi_{|\alpha}$. A strip on the surface is then the one-parameter set $x^\alpha(\lambda)$, $p_\alpha(\lambda)$ such that $\varphi(x^\alpha(\lambda)) \equiv 0$ and $p_\alpha(\lambda) = \varphi_{|\alpha}(x^\beta(\lambda))$. Geometrically, a strip is a one-parameter set of surface elements which are laid out along a curve $x^\alpha(\lambda)$ on S.

There are many possible strips on a given surface. Each curve $x^\alpha(\lambda)$ on S gives rise to such a strip. Among all those strips there is the distinguished set of characteristic strips which can be determined by means of a system of ordinary differential equations without knowledge of the surface S. Thus, instead of solving the partial differential equation (8.26), we may build up the solution surface S by its characteristic strips, which can be obtained as the solution of ordinary differential equations. The theory of ordinary differential equations is considered more elementary than that of partial differential equations. Thus the reduction of the partial differential equation (8.26) to the theory of characteristic strips represents a mathematical simplification.

However, the idea of characteristic strips is also of great geometrical and physical significance. Indeed, we shall show that it follows from the general theory that two integral surfaces φ = const of (8.26) which have a surface element in common, i.e., are tangent to each other at one point, have the entire characteristic strip through that point in common. Hence different integral surfaces of our partial differential equations have characteristic strips in common; this property can be taken as the definition of the characteristic strips. Thus they are the elementary building blocks of the integral surfaces. If we interpret the characteristic hypersurfaces of a partial differential system as the wave fronts of perturbations or signals, their characteristics in turn describe the propagation of localized perturbations along curves, i.e., rays of the propagating phenomenon. The characteristic curves on the characteristic hypersurfaces of the original system of partial differential equations are called the *bicharacteristics* of the original system.

To determine the characteristic strips on the integral surface S we proceed as follows: We assume, for the time being, that the solution $\varphi(x^\alpha)$ of (8.26) is known and define the curves $x^\alpha(\lambda)$ by means of the system of ordinary first-order differential equations

$$\dot{x}^\alpha(\lambda) = \frac{\partial H}{\partial \varphi_{|\alpha}} = 2g^{\alpha\beta}\varphi_{|\beta} \tag{8.27}$$

whose right side depends only on the $x^\alpha(\lambda)$. The integral curves of (8.27) will lie on the surface S, that is, satisfy $\varphi(x^\alpha)$ = const. Indeed, we have, by virtue of (8.26) and (8.27),

$$\frac{d}{d\lambda}\varphi(x^\alpha(\lambda)) = \varphi_{|\alpha}\dot{x}^\alpha = 2g^{\alpha\beta}\varphi_{|\alpha}\varphi_{|\beta} = 0 \tag{8.28}$$

The curves $x^\alpha(\lambda)$ defined by (8.27) now give rise to the strips $x^\alpha(\lambda)$, $p_\alpha(\lambda)$ on S, where

$$p_\alpha(\lambda) = \varphi_{|\alpha}(x^\mu(\lambda)) \tag{8.29}$$

We compute from (8.29) and (8.27)

$$\dot{p}_\alpha(\lambda) = \varphi_{|\alpha|\mu}\dot{x}^\mu = \frac{\partial H}{\partial \varphi_{|\mu}}\varphi_{|\alpha|\mu} \tag{8.30}$$

On the other hand, we may differentiate the identity (8.26) with respect to x^α and find

$$\frac{\partial H}{\partial x^\alpha} + \frac{\partial H}{\partial \varphi_{|\mu}}\varphi_{|\alpha|\mu} = 0 \tag{8.31}$$

Thus (8.30) simplifies to

$$\dot{p}_\alpha(\lambda) = -\frac{\partial H}{\partial x^\alpha} \tag{8.32}$$

At this point we can drop the assumption that the solution $\varphi(x^\alpha)$ of (8.26) is known. For (8.27) and (8.32) form a consistent ordinary differential system

$$\dot{x}^\alpha(\lambda) = \frac{\partial H(x^\alpha,p_\alpha)}{\partial p_\alpha} \qquad \dot{p}_\alpha(\lambda) = -\frac{\partial H(x^\alpha,p_\alpha)}{\partial x^\alpha} \tag{8.33}$$

which depends only on the known function

$$H(x^\alpha,p_\alpha) = g^{\alpha\beta}p_\alpha p_\beta \tag{8.34}$$

and can be integrated without knowledge of the solution $\varphi(x^\alpha)$ of the partial differential equation (8.26). The set of values $x^\alpha(\lambda)$, $p_\alpha(\lambda)$ is a characteristic strip on S; (8.33) is the equation system for such a characteristic strip. S can be built up from a manifold of such characteristic strips, depending on a sufficient number of parameters.

A characteristic strip is determined by the initial values $x^\alpha(0)$, $p_\alpha(0)$ and the system (8.33). Hence, if a strip has one element $x^\alpha(0)$, $p_\alpha(0)$ in common with a surface S, it will lie in it for all λ values, and if two surfaces S have a surface element in common, they must share the entire characteristic strip through it in common also, as we asserted at the beginning of this section.

In order to determine the bicharacteristics from the differential system (8.33), we draw on the well-known formalism of Hamiltonian mechanics. We may interpret the $x^\alpha(\lambda)$ and $p_\alpha(\lambda)$ as conjugate canonical variables in a mechanical system. We introduce the Lagrange function

$$L(x^\alpha,\dot{x}^\alpha) = \dot{x}^\alpha p_\alpha - H(x^\alpha,p_\alpha) \tag{8.35}$$

in which the p_α have been eliminated through the system of implicit equations

$$\dot{x}^\alpha = \frac{\partial H(x^\mu,p_\mu)}{\partial p_\alpha} \tag{8.36}$$

We then have the identities

$$\frac{\partial L}{\partial \dot{x}^\alpha} = p_\alpha \qquad \frac{\partial L}{\partial x^\alpha} = -\frac{\partial H}{\partial x^\alpha} \tag{8.37}$$

which are an immediate consequence of (8.35) and (8.36). By virtue of (8.37) the system (8.33) goes over into the Lagrangian form

$$\frac{d}{d\lambda}\left(\frac{\partial L}{\partial \dot{x}^\alpha}\right) = \frac{\partial L}{\partial x^\alpha} \tag{8.38}$$

Thus, we may globally characterize the bicharacteristics of our problem through a variational problem for which Eqs. (8.38) are the Euler-Lagrange equations. We find: The bicharacteristics of the Einstein equations are the extremals of the variational problem

$$\delta \int L\, d\lambda = 0 \tag{8.39}$$

with the Lagrange function L corresponding to the Hamiltonian (8.34). We find

$$L = \dot{x}^\alpha p_\alpha - g^{\alpha\beta}p_\alpha p_\beta \tag{8.40}$$

where we have to eliminate the p_α through the linear system

$$\dot{x}^\alpha = 2g^{\alpha\beta}p_\beta \tag{8.41}$$

This leads to

$$p_\alpha = \tfrac{1}{2}g_{\alpha\beta}\dot{x}^\beta \tag{8.42}$$

Furthermore, from the definition (8.34) of H in terms of p_α and the above, we obtain

$$H(x^\alpha,p_\alpha) = g^{\alpha\beta}p_\alpha p_\beta = \tfrac{1}{2}\dot{x}^\alpha p_\alpha = \tfrac{1}{4}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta \tag{8.43}$$

Therefore L is given as a function of x^α and $\dot{x}^\alpha$ by

$$L = \tfrac{1}{2}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta - \tfrac{1}{4}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = \tfrac{1}{4}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = H \tag{8.44}$$

Thus the bicharacteristics are simply the geodesics of our metric

$$\delta \int g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta\, d\lambda = 0 \tag{8.45}$$

which must, moreover, satisfy the side condition

$$g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = 0 \tag{8.46}$$

as a consequence of the differential equation (8.26), $H = 0$, and (8.44). Thus *the bicharacteristics of the Einstein equations are the null geodesics.*

The bicharacteristics of the Einstein equations for the metric tensor $g_{\alpha\beta}$ represent the lines along which perturbations of the metric can propagate. Since the metric determines the gravitational fields, we may say that gravitational perturbations spread along null geodesics. These are the same curves along which electromagnetic effects are propagated. Thus gravitational and electromagnetic phenomena have similar propagation properties, and the importance of the null geodesics for all wave phenomena is evident.

Our result on gravitational rays illustrates how a purely mathematical analysis of the structure of a differential system may lead to important physical insights and very significant results.

8.6 Uniqueness Problem for the Einstein Equations

As was mentioned in Sec. 8.3, Mme. Fourès-Bruhat has given an existence and uniqueness proof for the solution of the field equations of general relativity theory under very general assumptions. We shall restrict ourselves to a much simpler situation. We shall deal with the case of a static universe which is entirely free of matter and shall prove that it must possess a constant Lorentz metric throughout, if we demand that it tend to a Lorentzian universe at infinity. Since the constant Lorentz metric is obviously a solution of the Einstein field equations $R_{\mu\nu} = 0$ for the empty universe, we are dealing with a uniqueness problem for these field equations. This particular uniqueness proof is due to Lichnerowicz (1955). Its significance for the physical aspects of general relativity lies in the fact that it displays clearly the importance of the boundary conditions at infinity, i.e., the behavior of the universe at large. It shows that the local solution is strongly influenced by the global behavior of the metric and explains the central role of cosmological theories in general relativity.

We start out by precisely defining what we understand by a static space-time manifold (see Sec. 6.1 and Prob. 6.6). Such a manifold is characterized by the fundamental metric form

$$ds^2 = \xi^2(dx^0)^2 + g_{ij}\,dx^i\,dx^j \tag{8.47}$$

where the coefficients ξ and g_{ij} depend only on the three-space variables x^i and not on the time variable x^0. Since we demand the usual signature of a relativistic space-time metric, we must assume that the form $g_{ij}\,dx^i\,dx^j$ is negative-definite. Such a static line element was already considered in the case of the Schwarzschild metric; however, at present we do not make any assumption on spherical symmetry in space.

One may visualize a static space-time manifold as built up of identical layers of three-dimensional space hypersurfaces which are all orthogonal to the time-coordinate lines. Since the metric is independent of the time coordinate, all these three-dimensional spaces are isometric with the three-dimensional metric tensor g_{ij}, and if we disregard the unessential time coordinate x^0, we may identify them all with the "base space" ($x^0 = 0$, x^i). In this base space, we now have the tensor g_{ij} and the scalar field $\xi(x^i)$, which are combined through the field equations $R_{\mu\nu} = 0$.

Our problem is to find a solution system $\{g_{ij}, \xi(x^i)\}$ of these differential equations which is twice continuously differentiable in the entire base space and tends to the system $\{-\delta_{ij},1\}$ for x^i becoming infinite. We wish to show that these boundary conditions have the only solution

$$g_{ij}(x^k) \equiv -\delta_{ij} \qquad \xi(x^i) \equiv 1 \tag{8.48}$$

In order to study this uniqueness problem, we observe that the coordinate x^0 plays a distinguished role in our formulas. We shall use this fact to simplify the Einstein field equations and to express them in terms of the tensors and Christoffel symbols of the base space of the x^i. In order to distinguish the Riemann tensor and Christoffel symbols of the three-dimensional base space from the analogous quantities of the space-time manifold, we shall denote the three-dimensional base-space quantities by an asterisk.

We start out with the form of the metric tensor and its inverse

$$g_{\alpha\beta} = \begin{pmatrix} \xi^2 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & g_{ij} & \\ 0 & & & \end{pmatrix} \qquad g^{\alpha\beta} = \begin{pmatrix} \xi^{-2} & 0 & 0 & 0 \\ 0 & & & \\ 0 & & g^{ij} & \\ 0 & & & \end{pmatrix} \tag{8.49}$$

Next we compute the four-dimensional Christoffel symbols

$$\left\{ {\alpha \atop \beta\ \gamma} \right\} = g^{\alpha\tau}[\beta\gamma,\tau] = \tfrac{1}{2}g^{\alpha\tau}(g_{\beta\tau|\gamma} + g_{\gamma\tau|\beta} - g_{\beta\gamma|\tau}) \tag{8.50}$$

For all spatial indices we have, clearly,

$$\left\{ {i \atop k\ l} \right\} = \left\{ {i \atop k\ l} \right\}^* \tag{8.51}$$

since all g_{ik} are time-independent and all g_{i0} vanish. We next have to consider the Christoffel symbols with at least one zero index. Because of (8.49), we have

$$\left\{ {0 \atop \beta\ \gamma} \right\} = \frac{1}{2\xi^2}(g_{\beta 0|\gamma} + g_{\gamma 0|\beta}) \tag{8.52}$$

which leads to the three cases

$$\left\{ {0 \atop i\ k} \right\} = 0 \qquad \left\{ {0 \atop i\ 0} \right\} = \frac{\xi_{|i}}{\xi} \qquad \left\{ {0 \atop 0\ 0} \right\} = 0 \tag{8.53}$$

Similarly,

$$\left\{ {\alpha \atop 0\ \gamma} \right\} = \tfrac{1}{2}g^{\alpha\tau}(g_{0\tau|\gamma} - g_{0\gamma|\tau}) \tag{8.54}$$

which gives the additional cases

$$\left\{ {i \atop 0\ k} \right\} = 0 \qquad \left\{ {i \atop 0\ 0} \right\} = -\tfrac{1}{2}g^{il}g_{00|l} = -g^{il}\xi_{|l}\xi \tag{8.55}$$

From these formulas we now compute the values of the contracted Riemann tensor:

$$R_{\alpha\beta} = \left\{ {\rho \atop \beta\ \rho} \right\}_{|\alpha} - \left\{ {\rho \atop \alpha\ \beta} \right\}_{|\rho} + \left\{ {\rho \atop \alpha\ \sigma} \right\}\left\{ {\sigma \atop \beta\ \rho} \right\} - \left\{ {\rho \atop \alpha\ \beta} \right\}\left\{ {\sigma \atop \sigma\ \rho} \right\} \tag{8.56}$$

By virtue of (3.11), we have

$$\left\{ {\rho \atop \beta\ \rho} \right\} = (\log \sqrt{-g})_{|\beta} = (\log \xi + \log (\sqrt{-g^*}))_{|\beta} \tag{8.57}$$

and hence, because of the independence of our metric on x^0,

$$\begin{Bmatrix} \rho \\ 0\ \ \rho \end{Bmatrix} = 0 \tag{8.58}$$

We conclude from (8.51), (8.53), and (8.55) that

$$\begin{aligned} R_{ik} &= (\log \xi + \log \sqrt{-g^*})_{|i|k} - \begin{Bmatrix} r \\ i\ \ k \end{Bmatrix}^*_{|r} + \begin{Bmatrix} r \\ i\ \ s \end{Bmatrix}^* \begin{Bmatrix} s \\ k\ \ r \end{Bmatrix}^* \\ &\quad + \begin{Bmatrix} 0 \\ i\ \ 0 \end{Bmatrix} \begin{Bmatrix} 0 \\ k\ \ 0 \end{Bmatrix} - \begin{Bmatrix} r \\ i\ \ k \end{Bmatrix}^* (\log \xi + \log \sqrt{-g^*})_{|r} \\ &= R^*_{ik} + \frac{1}{\xi}\left(\xi_{|i|k} - \begin{Bmatrix} r \\ i\ \ k \end{Bmatrix}^* \xi_{|r}\right) \end{aligned} \tag{8.59}$$

To simplify the above relation we introduce the covariant vector in three-space,

$$\xi_i = \xi_{|i} \tag{8.60}$$

and its covariant derivative in the metric of the base space,

$$\xi_{i\|k} = \xi_{|i|k} - \begin{Bmatrix} r \\ i\ \ k \end{Bmatrix}^* \xi_r \tag{8.61}$$

In this notation we may express

$$R_{ik} = R^*_{ik} + \frac{1}{\xi}\,\xi_{i\|k} \tag{8.62}$$

as the relation between the contracted Riemann tensors in the two metrics.

Similarly, by (8.58), (8.53), and (8.55),

$$R_{i0} = 0 \tag{8.63}$$

Finally, we can write R_{00} by means of the above identities in the form

$$\begin{aligned} R_{00} &= -\begin{Bmatrix} r \\ 0\ \ 0 \end{Bmatrix}_{|r} + 2\begin{Bmatrix} 0 \\ 0\ \ r \end{Bmatrix}\begin{Bmatrix} r \\ 0\ \ 0 \end{Bmatrix} \\ &\quad - \begin{Bmatrix} r \\ 0\ \ 0 \end{Bmatrix}((\log \xi)_{|r} + (\log \sqrt{-g^*})_{|r}) \end{aligned} \tag{8.64}$$

We introduce the contravariant form of the vector ξ_i in the base space

$$\xi^i = g^{il}\xi_{|l} \tag{8.65}$$

Then, by (8.53) and (8.55), we can transform (8.64) into

$$\begin{aligned} R_{00} &= (\xi\xi^r)_{|r} - \xi_r\xi^r + \xi\xi^r(\log \sqrt{-g^*})_{|r} \\ &= \xi(\xi^r{}_{|r} + \xi^r(\log \sqrt{-g^*})_{|r}) = \xi\,\nabla^{*2}\xi \end{aligned} \tag{8.66}$$

Here we have used the identity (3.12),

$$\xi^r{}_{\|r} = \operatorname{div}^* \xi^r = \frac{1}{\sqrt{-g^*}}(\xi^r\sqrt{-g^*})_{|r} = \xi^r{}_{|r} + \xi^r(\log \sqrt{-g^*})_{|r} \tag{8.67}$$

and the definition of the Laplace operator ∇^{*2} in the base space in terms of the metric g_{ij}:

$$\nabla^{*2}\xi = \operatorname{div}^* \operatorname{grad} \xi = (g^{rs}\xi_{|s})_{\|r} \tag{8.68}$$

vgl. (3.14)

Thus the Einstein equations reduce to the following differential system in three-dimensional space:

$$\nabla^{*2}\xi = 0 \tag{8.69}$$

and

$$R^*_{ik} + \frac{1}{\xi}\,\xi_{i\|k} = 0 \tag{8.70}$$

We assume, of course, that $\xi(x^i) \neq 0$ since we wish to deal with a regular metric in space-time.

The procedure of solving the equation system (8.69) and (8.70) is now obvious. We shall first solve the Laplace equation (8.69) for $\xi(x^i)$, with the requirement that ξ be twice continuously differentiable in space and tend uniformly to 1 at infinity. Having determined $\xi(x^i)$ from these conditions, we insert it into (8.70) and determine g_{ij} from these equations.

If ∇^* were the ordinary Laplace operator, it would be evident that $\xi(x^i) \equiv 1$. Indeed, the solutions of the classical Laplace equation, the so-called harmonic functions, satisfy the maximum-minimum principle; i.e., in the neighborhood of each point P of regularity, there are points where the function takes values larger and smaller than at that point P itself. It follows that in each domain of regularity, the maximum and the minimum of the function are attained on the boundary of the domain. Since we assume that our solution is regular at every finite point and tends to the value 1 at infinity, it is clear that the maximum and the minimum of the solution must both be 1 and hence $\xi(x^i) \equiv 1$. We shall give in the next section a proof due to Hopf (1927), which extends the maximum-minimum principle to the generalized Laplacian needed here. Hence we may conclude from this principle that the differential equation (8.69) and the boundary conditions at infinity imply

$$\xi(x^i) \equiv 1 \tag{8.71}$$

If we insert the value (8.71) for ξ into (8.70), we find

$$R^*_{ik} = 0 \tag{8.72}$$

It is now a remarkable fact, which is of interest in its own right, that a three-dimensional space whose contracted Riemann tensor R^*_{ik} vanishes is a flat space; that is, its full Riemann tensor R^*_{iklm} is identically zero. Indeed, because of the numerous antisymmetries of the Riemann tensor, there are only a few nonzero independent components of the full Riemann tensor in three-space:

$$\begin{matrix} R^*_{12\,12} & R^*_{12\,13} & R^*_{12\,23} \\ R^*_{13\,13} & R^*_{13\,23} & R^*_{23\,23} \end{matrix} \tag{8.73}$$

Thus the number of independent components of the full Riemann tensor is precisely equal to six, which is also the number of the independent components of the symmetric contracted Riemann tensor. On the other hand, we can express by

$$R^*_{ik} = g^{lm} R^*_{milk} \tag{8.74}$$

the R^*_{ik} components as linear combinations of the six components of R^*_{milk}.

It is easily verified that the determinant of the linear transformation (8.74) is in general nonzero and that, conversely, the R^*_{milk} can be expressed linearly in terms of the R^*_{ik}. Thus Eqs. (8.72) imply indeed that R^*_{milk} vanishes identically. As we showed in Sec. 5.2, this implies that the base space is pseudo-Euclidean, and since we demand at infinity the limit condition $g_{ij} \to -\delta_{ij}$, we see that (8.48) is indeed fulfilled throughout the entire base space.

Thus, under the assumption of the maximum-minimum principle for the generalized Laplace equation (8.69), we have proved the uniqueness

of the Lorentz metric for an empty universe which becomes Lorentzian at spatial infinity.

8.7 The Maximum Principle for the Generalized Laplace Equation

We give, for the sake of completeness, the Hopf argument for the maximum-minimum principle for equations of the form

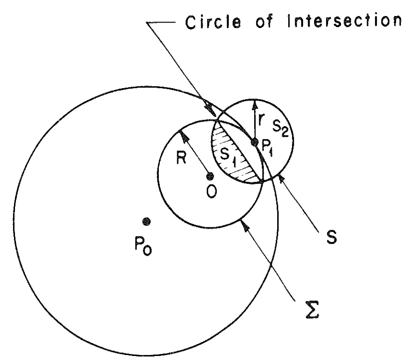
$$L[\xi] = g^{ik} \frac{\partial^2 \xi}{\partial x^i \, \partial x^k} + A^i \frac{\partial \xi}{\partial x^i} = 0 \tag{8.75}$$

with a negative-definite coefficient matrix g^{ik}. Clearly, our particular Laplace equation (8.69) is of this standard form. The proof is quite simple, and the result is of sufficient importance for many problems in applied mathematics and mathematical physics to justify its inclusion in this book.

Let us suppose that $\xi(x^i)$ is a solution of (8.75), defined in the entire space; assume it takes its minimum, say m, at some finite set of points. We select an arbitrary point P_0, where $\xi > m$, and consider that sphere around P_0 inside of which $\xi > m$, but on the surface of which the equation $\xi = m$ is fulfilled at one or more points. Such a sphere surely exists by our assumption. We next construct a sphere Σ which lies entirely inside this first sphere but which is tangent to it at a point P_1 at which $\xi = m$. Thus we can assert that inside and on the boundary of Σ, it is true that $\xi > m$, except for the one distinguished point P_1 at which $\xi = m$. The sphere Σ may furthermore be supposed centered at the origin and of radius R (Fig. 8.1).

We next consider one more sphere S with center P_1 and radius $r < R$. It intersects the sphere Σ, and its surface will be divided by Σ into two

Fig. 8.1



parts. We denote by S_1 that part of the surface S which is contained in Σ, including the curve of intersection (a circle), and by S_2 the open-surface region of S which is completely outside of Σ. Since $\xi(x^i)$ is continuous in the closed domain S_1, it has there a minimum and, by construction, this minimum is strictly greater than m. On S_2 we can only assert $\xi \geq m$, by our basic assumption. Thus there exists a positive number δ such that

$$\begin{aligned} &\xi(x^i) \geq m + \delta \qquad \text{on } S_1 \qquad \delta > 0 \\ &\xi(x^i) \geq m \qquad \text{on } S_2 \end{aligned} \tag{8.76}$$

Following Hopf, we introduce the function

$$h(x^i) = e^{-\alpha R^2} - e^{-\alpha r^2} \qquad \alpha > 0 \qquad r^2 = x^i x_i \tag{8.77}$$

where we define, for notational convenience, $x_i = \delta_{ij}x^j = x^j$. On Σ we have $h(x^i) = 0$, and everywhere

$$-1 < h(x^i) < 1 \tag{8.78}$$

A straightforward calculation yields

$$L[h] = e^{-\alpha r^2}(-4\alpha^2 g^{ik}x_i x_k + 2\alpha g^{ik}\delta_{ik} + 2\alpha A^i x_i) \tag{8.79}$$

Since g^{ik} is negative-definite, it is clear that we can choose α sufficiently large so that $L[h] > 0$ inside and on S. Since L is a linear operator on ξ, it follows also that, by virtue of (8.75),

$$L[\xi + \lambda h] > 0 \qquad \text{in and on } S \tag{8.80}$$

for all values $\lambda > 0$. Let us choose $0 < \lambda < \delta$, where δ is defined preceding (8.76).

We recall the bounds (8.78) for $h(x^i)$, and the fact that

$$\begin{aligned} &h(x^i) \leq 0 \qquad \text{on } S_1 \\ &h(x^i) > 0 \qquad \text{on } S_2 \end{aligned} \tag{8.81}$$

and also the inequalities (8.76). We conclude that, since $0 < \lambda < \delta$,

$$\begin{aligned} &\xi + \lambda h > m \qquad \text{on } S_1 \\ &\xi + \lambda h > m \qquad \text{on } S_2 \\ &\xi + \lambda h = m \qquad \text{at } P_1 \end{aligned} \tag{8.82}$$

Thus $\xi + \lambda h$ is a nonconstant, twice continuously differentiable function which must have a minimum inside of S, say at P_2. Hence, at P_2, the necessary minimum condition must hold:

$$\frac{\partial}{\partial x^i}(\xi + \lambda h)\bigg|_{P_2} = 0 \tag{8.83}$$

At P_2 the differential operator $L[\xi + \lambda h]$ becomes very simple, since all first derivatives of the argument function vanish. By (8.80), we can assert that

$$L[\xi + \lambda h] = g^{ik}\frac{\partial^2}{\partial x^i\,\partial x^k}(\xi + \lambda h)\bigg|_{P_2} > 0 \tag{8.84}$$

On the other hand, we know that g^{ik} is a negative-definite matrix, while the well-known necessary condition for the minimum of $\xi + \lambda h$ at P_2 implies

$$Q(t^i) = \frac{\partial^2}{\partial x^i\,\partial x^k}(\xi + \lambda h)\bigg|_{P_2} t^i t^k \geq 0 \tag{8.85}$$

for all real values t^i. We can bring the symmetric quadratic form $Q(t^i)$ onto principal axes; i.e., there exist linear forms

$$\tau^j = \alpha^j{}_i t^i \tag{8.86}$$

such that $Q(t^i)$ takes the form

$$Q(t^i) = \sum_{j=1}^{3} \lambda_j (\tau^j)^2 = \sum_{j=1}^{3} \lambda_j \alpha^j{}_i \alpha^j{}_k t^i t^k \tag{8.87}$$

with nonnegative eigenvalues λ_j. Hence, comparing the coefficients of $t^i t^k$ in (8.85) and (8.87), we can assert

$$\frac{\partial^2}{\partial x^i\,\partial x^k}(\xi + \lambda h)\bigg|_{P_2} = \sum_{j=1}^{3} \lambda_j \alpha^j{}_i \alpha^j{}_k \tag{8.88}$$

and (8.84) becomes

$$\sum_{j=1}^{3} \lambda_j (g^{ik}\alpha^j{}_i \alpha^j{}_k) > 0 \tag{8.89}$$

This contradicts the fact that g^{ik} is negative-definite and the λ_j are nonnegative. Thus our assumption regarding a finite minimum point leads to an obvious contradiction. If we replace ξ by $-\xi$, we see at once that a finite maximum point is likewise excluded. Thus the proof on the nonexistence of a finite maximum or minimum within the domain of definition is complete.

Exercises

8.1 In special relativity Maxwell's equations for free space may be written in terms of the four-vector potential A^μ as

$$\Box^2 A^\mu = g^{\nu\alpha} A^\mu{}_{|\nu|\alpha} = 0$$

if we impose the Lorentz condition $A^\mu{}_{|\mu} = 0$ (see Exercise 4.4). Formulate the initial-value problem for this system.

8.2 (*continued*) Prove that

$$\nabla^2 A^0 = \frac{\partial}{\partial t} \nabla \cdot \mathbf{A}$$

so that the component A^0 is not the solution of a typical Cauchy problem of second-order differential equations.

8.3 Must all partial differential equations or systems of partial differential equations possess characteristic surfaces? If an equation or system of equations does not possess characteristic surfaces, what sort of physical process could it describe, and what sort of physical process could it not describe?

8.4 The Schrödinger equation, which describes the motion of a free particle in nonrelativistic quantum theory, is of the form

$$i\lambda \frac{\partial \psi}{\partial t} + \nabla^2 \psi = 0 \qquad \psi = \psi(\mathbf{x},t)$$

where λ is a real constant. [$|\psi(\mathbf{x},t)|^2$ is interpreted as the relative probability of the particle's being at position $\mathbf{x}$ at time t.] Discuss the Cauchy problem and the existence of characteristic surfaces for this equation.

8.5 The Klein-Gordon equation replaces the Schrödinger equation in the relativistic quantum theory of mesons. It has the form, in flat space,

$$\Box^2 \psi + \tau^2 \psi = 0$$

where τ is a real constant. Discuss the Cauchy problem and the existence of characteristic surfaces for this equation.

8.6 How would the study of differential equations change if we considered equations of higher than second order? What Cauchy data would be needed? Would characteristic surfaces occur?

8.7 Obtain the characteristic surfaces of Maxwell's equations in flat space. You may wish to consult Chap. 4 for one method of doing this. Obtain also the bicharacteristics and show that they represent null geodesics, physically interpreted as light rays. Illustrate this in a picture with the z coordinate suppressed, analogous to Fig. 7.1.

8.8 Draw analogies between the mathematical structure of Maxwell's equations in flat space and the Einstein equations, and between the characteristics and bicharacteristics of the two systems of equations.

8.9 Show how the uniqueness theorem of Sec. 8.6 breaks down if the function ξ is allowed to have a singularity.

Problems

8.1 Investigate the characteristics and bicharacteristics of the Maxwell equations in a general Riemann space.

8.2 In the cosmological problem we do not deal with Euclidean boundary conditions at infinity (see Chap. 12). One of the possible forms taken by the metric is, in Cartesian coordinates,

$$ds^2 = c^2\, dt^2 - R(t)^2[dx^2 + dy^2 + dz^2]$$

Analyze the uniqueness problem for this metric form instead of that discussed in the text, (8.47).

Bibliography

Bateman, H. (1932): "Partial Differential Equations," Cambridge, England.
Courant, R., and D. Hilbert (1962): "Methods of Mathematical Physics," vol. 2, "Partial Differential Equations," New York.
Darmois, G. (1927): "Les équations de la gravitation einsteinienne," Paris.
Duff, G. F. D. (1956): "Partial Differential Equations," Toronto.
Fourès-Bruhat, Y. (1952): Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires, *Acta Math.*, **88**:141–225.
Fourès-Bruhat, Y. (1962): The Cauchy Problem, in L. Witten (ed.), "Gravitation: An Introduction to Current Research," New York, pp. 130–168.

Hadamard, J. (1932): "Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques," Paris.
Hilbert, D. (1915): Die Grundlagen der Physik, I, *Göttinger Nachr.*, pp. 395–407.
Hopf, E. (1927): Elementare Bemerkungen über die Lösungen partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 147–152.
Lichnerowicz, A. (1939): "Problèmes globaux en mécanique relativiste," Paris.
Lichnerowicz, A. (1955): "Théories relativistes de la gravitation et de l'électromagnétisme," Paris.
Taub, A. H. (1957): Singular Hypersurfaces in General Relativity, *Illinois J. Math.*, **1**:370–388.

CHAPTER

9

The Linearized Field Equations

The field equations for free space which we developed in Chap. 5 are clearly not linear, so the superposition principle does *not* hold; that is, if $g_{\alpha\beta}$ and $g^*_{\alpha\beta}$ are solutions to the field equations, then a linear combination of the two is *not* necessarily a solution. The physical reason for this is easily understood, for the gravitational field of a body can do work and must therefore contain energy. Since it possesses energy, it must possess effective mass, and thereby create a *further* gravitational field. That is, the field *itself* can serve as part of *its own source*. Because of this feedback effect the gravitational field produced by two bodies is not a simple sum of the separate fields of the two bodies, but involves the detailed structure of the interacting fields.

In this chapter we shall develop a set of approximate linear equations in which the feedback effect of the gravitational field as its own source is ignored. In such an approximate theory gravitational effects are considered to be simply additive, as they are in the classical gravitational theory. Furthermore, because of their linearity, these equations will possess the virtue of being mathematically simpler than the exact gravitational field equations of Chap. 5.

9.1 Linearization of the Field Equations

In order that we may ignore the feedback effect of the gravitational field as part of its own source, it is clear that we must assume that the field is weak. Thus we shall deal throughout this chapter with a metric tensor which differs only slightly from the flat-space metric tensor. Accordingly, we may write the metric tensor as the flat-space Lorentz metric tensor $\eta_{\alpha\beta}$ plus a perturbation term $\epsilon\gamma_{\alpha\beta}$,

$$g_{\alpha\beta} = \eta_{\alpha\beta} + \epsilon\gamma_{\alpha\beta} \tag{9.1}$$

and consider only first-order terms in the parameter ϵ as significant in all equations.

It will be convenient in this chapter to use the Minkowski coordinates ict, x, y, and z in place of the usual coordinates of special relativity. In this coordinate system the Lorentz metric tensor has the simple form

$$\eta_{\alpha\beta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{9.2}$$

that is, $\eta_{\alpha\beta}$ is simply the negative of the Kronecker delta $\delta_{\alpha\beta}$.

Using the Minkowski coordinates and the form given in (9.2) for $\eta_{\alpha\beta}$, let us consider the free-space gravitational field equations (5.119):

$$0 = R_{\eta\lambda} = \left\{ {\beta \atop \beta\ \eta} \right\}_{|\lambda} - \left\{ {\beta \atop \lambda\ \eta} \right\}_{|\beta} + \left\{ {\beta \atop \tau\ \lambda} \right\}\left\{ {\tau \atop \beta\ \eta} \right\} - \left\{ {\beta \atop \tau\ \beta} \right\}\left\{ {\tau \atop \lambda\ \eta} \right\} \tag{9.3}$$

Since the Christoffel symbols are homogeneous and linear in the first derivatives of the metric tensor, and since the $\eta_{\alpha\beta}$ have vanishing derivatives, each of the first two terms of (9.3) will contain a single factor of the parameter ϵ. The second two terms, however, are products of Christoffel symbols and therefore will contain factors of ϵ^2; by our approximation scheme, these second two terms are to be ignored. Equation (9.3) to first order in ϵ is then

$$\left\{ {\beta \atop \beta\ \eta} \right\}_{|\lambda} - \left\{ {\beta \atop \lambda\ \eta} \right\}_{|\beta} = 0 \tag{9.4}$$

The first contracted Christoffel symbol may be simplified; from (3.11) we have

$$\left\{ {\beta \atop \beta\ \eta} \right\} = \tfrac{1}{2}(\log |g|)_{|\eta} \tag{9.5}$$

so we can put (9.4) in the form

$$\tfrac{1}{2}(\log |g|)_{|\eta|\lambda} - \left\{ {\beta \atop \lambda\ \eta} \right\}_{|\beta} = 0 \tag{9.6}$$

The function $|g|$, the absolute value of the metric-tensor determinant, can be written, using (9.1), as

$$|g| = \|-g_{\alpha\beta}\| = \begin{Vmatrix} 1-\epsilon\gamma_{00} & -\epsilon\gamma_{01} & -\epsilon\gamma_{02} & -\epsilon\gamma_{03} \\ -\epsilon\gamma_{10} & 1-\epsilon\gamma_{11} & -\epsilon\gamma_{12} & -\epsilon\gamma_{13} \\ -\epsilon\gamma_{20} & -\epsilon\gamma_{21} & 1-\epsilon\gamma_{22} & -\epsilon\gamma_{23} \\ -\epsilon\gamma_{30} & -\epsilon\gamma_{31} & -\epsilon\gamma_{32} & 1-\epsilon\gamma_{33} \end{Vmatrix} \tag{9.7}$$

But the only term of first order in ϵ which occurs in this determinant comes from the product of the diagonal elements, so to first order in ϵ, we have

$$\begin{aligned} |g| &= (1-\epsilon\gamma_{00})(1-\epsilon\gamma_{11})(1-\epsilon\gamma_{22})(1-\epsilon\gamma_{33}) \\ &= 1 - \epsilon(\gamma_{00}+\gamma_{11}+\gamma_{22}+\gamma_{33}) \\ &= 1 - \epsilon \operatorname{Tr} \gamma \end{aligned} \tag{9.8}$$

Expanding $\log |g|$ in a Taylor series to first order in ϵ, we obtain

$$\log |g| = \log (1 - \epsilon \operatorname{Tr} \gamma) = -\epsilon \operatorname{Tr} \gamma \tag{9.9}$$

Thus the first term of the linearized equations (9.6) may be written as

$$\tfrac{1}{2}(\log |g|)_{|\eta|\gamma} = -\tfrac{1}{2}\epsilon(\operatorname{Tr} \gamma)_{|\eta|\gamma} = -\tfrac{1}{2}\epsilon \sum_{\beta=0}^{3} \gamma_{\beta\beta|\eta|\gamma} \tag{9.10}$$

The second Christoffel symbol which appears in (9.4) may be written out as

$$\left\{ {\beta \atop \lambda\ \eta} \right\} = \frac{g^{\beta\mu}}{2}[\lambda\eta,\mu] = \frac{\eta^{\beta\mu} + \epsilon\gamma^{\beta\mu}}{2}(\epsilon\gamma_{\mu\lambda|\eta} + \epsilon\gamma_{\mu\eta|\lambda} - \epsilon\gamma_{\lambda\eta|\mu}) \tag{9.11}$$

Using the explicit form given for $\eta_{\alpha\beta}$ in (9.2), we can write this to first order in ϵ as

$$\left\{ {\beta \atop \lambda\ \eta} \right\} = -\tfrac{1}{2}(\epsilon\gamma_{\beta\lambda|\eta} + \epsilon\gamma_{\beta\eta|\lambda} - \epsilon\gamma_{\lambda\eta|\beta}) \tag{9.12}$$

(Note that we are *not* raising or lowering indices in the above.) Hence, to first order in ϵ, we obtain

$$\left\{ {\beta \atop \lambda\ \eta} \right\}_{|\beta} = -\sum_{\beta=0}^{3} \tfrac{1}{2}(\epsilon\gamma_{\beta\lambda|\eta} + \epsilon\gamma_{\beta\eta|\lambda} - \epsilon\gamma_{\lambda\eta|\beta})_{|\beta} \tag{9.13}$$

By use of (9.10) and (9.13), the linearized equations (9.3) can now be written completely in terms of the perturbation $\gamma_{\alpha\beta}$ as

$$\sum_{\beta=0}^{3} \gamma_{\beta\beta|\eta|\lambda} - \sum_{\beta=0}^{3} (\gamma_{\beta\eta|\lambda|\beta} + \gamma_{\lambda\beta|\eta|\beta} - \gamma_{\eta\lambda|\beta|\beta}) = 0 \tag{9.14}$$

We have now obtained a system of 10 partial differential equations for the 10 unknown components of the symmetric $\gamma_{\alpha\beta}$ perturbation term. In the following sections we shall investigate and simplify these equations.

We should note at this point that, since we delete terms of second and higher order in the parameter ϵ, we no longer have a covariant theory; that is, the linearized equations are not covariant and the solution $\gamma_{\alpha\beta}$ is therefore not necessarily a tensor. However, it is easily shown that we may treat the approximate metric tensor $g_{\alpha\beta}^{(A)} = \eta_{\alpha\beta} + \epsilon\gamma_{\alpha\beta}$ as a true tensor under coordinate transformations consistently to order ϵ. To show this, let us form the difference between the exact metric tensor $g_{\alpha\beta}$, which is a solution of the complete field equations (9.3), and the corresponding approximate solution $g_{\alpha\beta}^{(A)}$ obtained from the linearized equations (9.14). According to our approximation scheme, this difference is of order ϵ^2. Thus we may write

$$g_{\alpha\beta} - g_{\alpha\beta}^{(A)} = O(\epsilon^2) \tag{9.15}$$

Let us now tentatively treat $g_{\alpha\beta}^{(A)}$ as a tensor and define in a barred coordinate system, which may or may not be Minkowskian,

$$\bar{g}_{\mu\nu}^{(A)} = \frac{\partial x^\alpha}{\partial \bar{x}^\mu}\frac{\partial x^\beta}{\partial \bar{x}^\nu} g_{\alpha\beta}^{(A)} \tag{9.16}$$

It follows that the difference between the "approximate metric tensor" so defined and the exact metric tensor in the new system is

$$\bar{g}_{\mu\nu} - \bar{g}_{\mu\nu}^{(A)} = \frac{\partial x^\alpha}{\partial \bar{x}^\mu}\frac{\partial x^\beta}{\partial \bar{x}^\nu} (g_{\alpha\beta} - g_{\alpha\beta}^{(A)}) = O(\epsilon^2) \tag{9.17}$$

Thus, by treating $g_{\alpha\beta}^{(A)}$ as a true tensor under coordinate transformation, we make an error at most of order ϵ^2. Since we shall carry out all our calculations with deliberate neglect of all terms containing the factor ϵ^2, we may treat $g_{\alpha\beta}^{(A)}$ consistently as an ordinary tensor. Furthermore, it is clear that, if the above transformation is to a new Minkowski frame where we can solve the linearized equations, then the approximate metric defined by (9.16) and the approximate metric obtained by solving the linearized equations in the new frame can differ at most by order ϵ^2.

9.2 The Time-independent and Spherically Symmetric Field

We possess in the linearized field equations (9.14) a convenient tool for studying the influence of weak perturbations on a flat-space metric. Clearly, the use of linear equations for the calculation of the metric tensor is closest to the approach of classical potential theory and allows the best comparison between classical and relativistic theory. Historically, the linearized theory was first used to investigate the static gravitational field with spherical symmetry; the experimentally detectable effects which we studied in Secs. 4.4, 6.3, and 6.5 were first obtained by Einstein in this manner in 1915. Shortly afterward, in 1916, Schwarzschild succeeded in obtaining an exact solution to the same problem, which we have already studied in Chap. 6. We shall now study the approximate solution of Einstein. The reason for doing so is twofold. First, we present Einstein's argument for historical interest, and second, we obtain experience in handling the system (9.14). It will become apparent that even in cases where an exact solution can be obtained, the linearized solution is valuable, in that it is more accessible to calculation and is often more open to physical interpretation.

As we discussed in Chap. 6, the meaning of a static gravitational field may be summed up in the form of its line element

$$ds^2 = g_{00}(dx^0)^2 + g_{ik}\,dx^i\,dx^k \tag{9.18}$$

where g_{00} and g_{ik} depend only on the space variables x^i. If we use Minkowski coordinates as before, we have

$$x_0 = ict \tag{9.19}$$

and hence we can express (9.18) as

$$ds^2 = -g_{00}c^2\,dt^2 + g_{ik}\,dx^i\,dx^k \tag{9.20}$$

with a negative definite g_{ik} matrix.

In order to carry out the linearization described in the preceding section, we introduce a function $a(x^i)$ defined by

$$g_{00} = -1 + \epsilon a \qquad a = \gamma_{00} \qquad \gamma_{0i} = 0 \tag{9.21}$$

and write g_{ik} in the form

$$g_{ik} = -\delta_{ik} + \epsilon\gamma_{ik} \tag{9.22}$$

With these definitions we can write the $\eta = \lambda = 0$ component of the linearized system (9.14),

$$\sum_{\beta=0}^{3} \gamma_{\beta\beta|0|0} - \sum_{\beta=0}^{3} (\gamma_{\beta 0|0|\beta} + \gamma_{0\beta|0|\beta} - \gamma_{00|\beta|\beta}) = 0 \tag{9.23}$$

in a simple form. Since all $\gamma_{\alpha\beta}$ are independent of x^0, the only surviving term of (9.23) is

$$\sum_{\beta=0}^{3} \gamma_{00|\beta|\beta} = \sum_{k=1}^{3} a_{|k|k} = 0 \tag{9.24}$$

In terms of the Laplacian operator ∇^2, this is simply the harmonic equation in three dimensions,

$$\nabla^2 a = \nabla^2 \gamma_{00} = 0 \tag{9.25}$$

Thus we have shown that γ_{00} of the perturbation matrix satisfies Laplace's equation and is therefore a harmonic function.

Let us digress for a moment and consider the implications of (9.25) with regard to the correspondence between general relativity theory and classical gravitational theory. In studying the geodesic equations of motion in Sec. 4.3, we found that they correspond to Newton's classical equations of motion in a potential field in the limit of low velocities and weak fields. However, in order to make the correspondence hold, we had to assume an asymptotic relation between the g_{00} component of the metric tensor and the classical potential φ, which describes the gravitational field; in terms of the coordinates of special relativity ct, x, y, and z, that relation was

$$g_{00} = 1 + \frac{2\varphi}{c^2} \qquad \text{(coordinates of special relativity)} \tag{9.26}$$

[Eq. (4.142)]. In terms of the Minkowski coordinates ict, x, y, and z, which we are now using, (9.26) becomes

$$g_{00} = -1 - \frac{2\varphi}{c^2} \qquad \text{(Minkowski coordinates)} \tag{9.27}$$

Using the above equation and the definition of the perturbation matrix $\gamma_{\mu\nu}$ in (9.1), we see that we must have approximately

$$\epsilon\gamma_{00} = \epsilon a = -\frac{2\varphi}{c^2} \tag{9.28}$$

This equation allows us to deduce a field equation for the classical potential φ on the basis of relativity theory. Indeed, from (9.25), we obtain

$$\nabla^2 \varphi = 0 \tag{9.29}$$

which is the same equation, Laplace's equation, that φ satisfies according to classical theory.

Let us note that the identification of geodesic motion in relativity theory with Newtonian motion in a potential field led to the correspondence (9.26) between metric and potential functions without the use of Einstein's field equations. We now see that Einstein's field equations in the linearized approximation are consistent with the correspondence, and furthermore reduce precisely to the correct classical equation in the limit considered above.

One other fact concerning the classical correspondence of general relativity is worthy of note at this point. It is well known in classical potential theory that if the harmonic function φ satisfies the boundary condition that it be zero at infinity, then it must be zero everywhere, unless Laplace's equation breaks down at some point in space or in some extended region of space. Physically, this means that there must be some point (a particle) or some region (an extended body) where $\nabla^2\varphi$ is nonzero, or else there can be no gravitational field. Clearly, the same result holds for the function $\gamma_{00} = a$ by virtue of Eq. (9.28).

The above result is analogous to the uniqueness theorem of Sec. 8.6. However, it is not a special case of that theorem, since we are now dealing with the linearized theory only.

Until now we have used only the assumption that the gravitational field is static. Let us now add the assumption of radial symmetry. As we discussed in Sec. 6.2, a radially symmetric static line element can be put into the isotropic form

$$ds^2 = -g_{00}c^2\,dt^2 - g_{11}(dx^2 + dy^2 + dz^2) \tag{9.30}$$

If we denote the usual radial distance by $r = \sqrt{x^2 + y^2 + z^2}$, we can, moreover, assert that, because of radial symmetry,

$$g_{00} = -1 + \epsilon a(r) \qquad g_{11} = -1 + \epsilon b(r) \tag{9.31}$$

and thus the perturbation matrix $\gamma_{\alpha\beta}$ assumes the simple form

$$\gamma_{\alpha\beta}(r) = \begin{pmatrix} a(r) & 0 & 0 & 0 \\ 0 & b(r) & 0 & 0 \\ 0 & 0 & b(r) & 0 \\ 0 & 0 & 0 & b(r) \end{pmatrix} \tag{9.32}$$

By the correspondence (9.28) we can easily obtain the function $a(r)$ explicitly; indeed, the classical potential of a spherically symmetric field (with a singularity at $r = 0$) is simply $\varphi = -\kappa M/r$, where M is the mass of the body at $r = 0$ and κ is the gravitational constant. Thus, from (9.28),

$$\epsilon a(r) = \frac{2\kappa M}{c^2 r} \tag{9.33}$$

We next wish to determine the remaining unknown function $b(r)$ by investigating the remaining linearized field equations for $\gamma_{11} = \gamma_{22} = \gamma_{33}$. We set $\eta = \lambda = j$ in the linearized equations (9.14) to obtain

$$\sum_{\beta=0}^{3} \gamma_{\beta\beta|j|j} = \sum_{\beta=0}^{3} (\gamma_{\beta j|j|\beta} + \gamma_{j\beta|j|\beta} - \gamma_{jj|\beta|\beta}) \tag{9.34}$$

In terms of $a(r)$ and $b(r)$, the left side of (9.34) becomes

$$\sum_{\beta=0}^{3} \gamma_{\beta\beta|j|j} = \gamma_{00|j|j} + \sum_{k=1}^{3} \gamma_{kk|j|j} = a_{|j|j} + 3b_{|j|j} \tag{9.35}$$

Since the matrix $\gamma_{\alpha\beta}$ is diagonal, symmetric in the indices α and β, and independent of time, the right side of (9.34) becomes

$$\begin{aligned}\sum_{\beta=0}^{3} (\gamma_{\beta j|j|\beta} + \gamma_{j\beta|j|\beta} - \gamma_{jj|\beta|\beta}) &= 2\gamma_{jj|j|j} - \sum_{k=1}^{3} \gamma_{jj|k|k} \\ &= 2b_{|j|j} - \sum_{k=1}^{3} b_{|k|k}\end{aligned} \tag{9.36}$$

Using Eqs. (9.35) and (9.36), we can rewrite the linearized equation (9.34) in the form

$$a_{|j|j} + 3b_{|j|j} = 2b_{|j|j} - \sum_{k=1}^{3} b_{|k|k} \tag{9.37}$$

In terms of the Laplacian operator ∇^2, this can be written as

$$a_{|j|j} + b_{|j|j} + \nabla^2 b = 0 \tag{9.38}$$

Now let us sum Eqs. (9.38) over values of j and recall that $\nabla^2 a = 0$. We find

$$\nabla^2 b = 0 \tag{9.39}$$

and hence (9.38) reduces to

$$(a + b)_{|j|j} = 0 \tag{9.40}$$

We conclude that $a + b$ is a linear function of the coordinates, and since it is zero at infinity, it is zero everywhere. Hence we have proved $a = -b$, and we therefore arrive at the following perturbation matrix:

$$\gamma_{\alpha\beta} = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & -a & 0 & 0 \\ 0 & 0 & -a & 0 \\ 0 & 0 & 0 & -a \end{pmatrix} \tag{9.41}$$

The approximate metric tensor in Minkowski coordinates follows from (9.41):

$$g_{\alpha\beta} = \begin{pmatrix} -1 + \epsilon a & 0 & 0 & 0 \\ 0 & -1 - \epsilon a & 0 & 0 \\ 0 & 0 & -1 - \epsilon a & 0 \\ 0 & 0 & 0 & -1 - \epsilon a \end{pmatrix} \tag{9.42}$$

The line element is, accordingly,

$$ds^2 = (1 - \epsilon a)c^2\, dt^2 - (1 + \epsilon a)(dx^2 + dy^2 + dz^2) \tag{9.43}$$

Using the classical correspondence relations (9.28) and (9.33), we can write this in the form

$$\begin{aligned} ds^2 &= \left(1 + \frac{2\varphi}{c^2}\right) c^2\, dt^2 - \left(1 - \frac{2\varphi}{c^2}\right)(dx^2 + dy^2 + dz^2) \\ &= \left(1 - \frac{2kM}{c^2 r}\right) c^2\, dt^2 - \left(1 + \frac{2kM}{c^2 r}\right) d\sigma^2 \end{aligned} \tag{9.44}$$

where $d\sigma^2 = dx^2 + dy^2 + dz^2$. This result agrees with the exact Schwarzschild solution in isotropic coordinates, which we studied in Chap. 6.

9.3 The Weyl Solutions to the Linearized Field Equations

In this section we shall obtain a particularly interesting class of solutions to the linearized equations which is due to Weyl (Weyl, 1918). The rest of the chapter will then be devoted to investigating the relation of the Weyl solutions to the structure of the linearized equations.

We return now to the general linearized equations (9.14) and define the four-dimensional D'Alembertian operator

$$\Box^2\gamma_{\eta\lambda} = -\sum_{\beta=0}^{3} \gamma_{\eta\lambda|\beta|\beta} \tag{9.45}$$

In terms of this operator the linearized field equations (9.14) can be written as

$$\Box^2\gamma_{\eta\lambda} + \sum_{\beta=0}^{3} \gamma_{\lambda\beta|\eta|\beta} + \sum_{\beta=0}^{3} \gamma_{\beta\eta|\lambda|\beta} - \sum_{\beta=0}^{3} \gamma_{\beta\beta|\eta|\lambda} = 0 \tag{9.46}$$

which can be rearranged into the symmetric form

$$\Box^2\gamma_{\eta\lambda} + \sum_{\beta=0}^{3} (\gamma_{\lambda\beta|\beta} - \tfrac{1}{2}\gamma_{\beta\beta|\lambda})_{|\eta} + \sum_{\beta=0}^{3} (\gamma_{\eta\beta|\beta} - \tfrac{1}{2}\gamma_{\beta\beta|\eta})_{|\lambda} = 0 \tag{9.46$'$}$$

If we define a four-component quantity τ_η as

$$\tau_\eta = -\sum_{\beta=0}^{3} (\gamma_{\eta\beta|\beta} - \tfrac{1}{2}\gamma_{\beta\beta|\eta}) \tag{9.47}$$

this can be written in more compact form as

$$\Box^2\gamma_{\eta\lambda} = \tau_{\lambda|\eta} + \tau_{\eta|\lambda} \tag{9.48}$$

which is completely equivalent to (9.14).

Following Weyl, let us now tentatively set τ_λ equal to the D'Alembertian of a four-component function φ_λ,

$$\tau_\lambda = \Box^2\varphi_\lambda \tag{9.49}$$

which can be obtained by solving the inhomogeneous wave equation. We substitute this in (9.48):

$$\Box^2\gamma_{\eta\lambda} = \Box^2\varphi_{\lambda|\eta} + \Box^2\varphi_{\eta|\lambda} \tag{9.50}$$

This leads us to investigate the possibility that a solution for the matrix $\gamma_{\eta\lambda}$ might be

$$\gamma_{\eta\lambda} = \varphi_{\lambda|\eta} + \varphi_{\eta|\lambda}, \tag{9.51}$$

where the φ_λ are four arbitrarily assigned functions. By substituting this back into (9.46), we can easily verify that it is indeed a solution. The left side of (9.46) becomes

$$\begin{aligned}
&\Box^2\gamma_{\eta\lambda} + \sum_{\beta=0}^{3} (\gamma_{\lambda\beta|\eta|\beta} + \gamma_{\beta\eta|\lambda|\beta} - \gamma_{\beta\beta|\eta|\lambda}) \\
&= \Box^2\gamma_{\eta\lambda} + \sum_{\beta=0}^{3} (\varphi_{\lambda|\beta|\eta|\beta} + \varphi_{\beta|\lambda|\eta|\beta} + \varphi_{\beta|\eta|\lambda|\beta} + \varphi_{\eta|\beta|\lambda|\beta} - \varphi_{\beta|\beta|\eta|\lambda} - \varphi_{\beta|\beta|\eta|\lambda}) \\
&= \Box^2\gamma_{\eta\lambda} - (\Box^2\varphi_{\lambda|\eta} + \Box^2\varphi_{\eta|\lambda})
\end{aligned} \tag{9.52}$$

By virtue of (9.50) this is zero, so the expression (9.51) is indeed a solution of the field equations.

Let us consider for a moment the above result. We began with the general linearized system (9.46) of 10 equations for the 10 independent elements of the symmetric matrix $\gamma_{\eta\lambda}$. By Weyl's *Ansatz* (9.51) we were able to generate a large subclass of solutions by using an arbitrary twice-differentiable set of four functions φ_λ. Solutions which belong to this subclass, i.e., have the form $\varphi_{\lambda|\eta} + \varphi_{\eta|\lambda}$, are termed Weyl solutions. We shall see that they form a very important subclass of solutions to the linearized equations.

Consider now an arbitrary solution of the linearized equations $\gamma_{\eta\lambda}$. Using this solution, we define a set of four associated τ_η functions as

$$\tau_\eta = -\sum_{\beta=0}^{3} (\gamma_{\eta\beta|\beta} - \tfrac{1}{2}\gamma_{\beta\beta|\eta}) \tag{9.53}$$

We can also define a set of associated φ_λ functions by the four equations

$$\Box^2\varphi_\lambda = \tau_\lambda \tag{9.54}$$

The resultant set of functions φ_λ, which we have generated from the original solution of the field equations $\gamma_{\eta\lambda}$, can now serve as a generating function for a new solution of the field equations of the Weyl form

$$\gamma_{\eta\lambda}^{(w)} = \varphi_{\eta|\lambda} + \varphi_{\lambda|\eta} \tag{9.55}$$

which, we note, has the same set of associated τ_η functions (9.53) as $\gamma_{\eta\lambda}$. This solution which we have generated will be termed the associated Weyl solution of $\gamma_{\eta\lambda}$. Thus, to every solution of the linearized equations, there corresponds a unique *associated* Weyl solution.

There is also another type of solution associated with the arbitrary solution $\gamma_{\eta\lambda}$ which will prove to be of interest; consider the difference between $\gamma_{\eta\lambda}$ and its associated Weyl solution $\gamma_{\eta\lambda}^{(w)}$:

$$\gamma_{\eta\lambda} - \gamma_{\eta\lambda}^{(w)} = \hat{\gamma}_{\eta\lambda} \tag{9.56}$$

Since the equations (9.14) are linear, this is indeed a solution, and we note, furthermore, that it is uniquely determined by $\gamma_{\eta\lambda}$.

Let us investigate this solution by first computing its D'Alembertian. From the definition of τ_η in (9.53) and the field equations in the form (9.48), the D'Alembertian of $\gamma_{\eta\lambda}$ is simply

$$\Box^2\gamma_{\eta\lambda} = \tau_{\eta|\lambda} + \tau_{\lambda|\eta} \tag{9.57}$$

Similarly, from the definitions of $\gamma_{\eta\lambda}^{(w)}$ in (9.55) and of φ_λ in (9.54), we obtain

$$\Box^2\gamma_{\eta\lambda}^{(w)} = \Box^2\varphi_{\eta|\lambda} + \Box^2\varphi_{\lambda|\eta} = \tau_{\eta|\lambda} + \tau_{\lambda|\eta} \tag{9.58}$$

Thus the solution $\gamma_{\eta\lambda}$ and its associated Weyl solution have the same D'Alembertian. The difference of these solutions $\hat{\gamma}_{\eta\lambda}$ therefore has a null D'Alembertian:

$$\Box^2\hat{\gamma}_{\eta\lambda} = 0 \tag{9.59}$$

This is a very important result, for Eq. (9.59) is the familiar wave equation of classical physics. It states that the 10-component disturbance which is represented by $\hat{\gamma}_{\eta\lambda}$ is *propagated with velocity c, the speed of light.*

We shall next obtain a set of equations which relates the components $\hat{\gamma}_{\eta\lambda}$ to each other. The set of functions $\hat{\tau}_\eta$ associated with $\hat{\gamma}_{\eta\lambda}$ and defined by an equation analogous to (9.53) obeys the equations

$$\Box^2\hat{\gamma}_{\eta\lambda} = \hat{\tau}_{\eta|\lambda} + \hat{\tau}_{\lambda|\eta} = 0 \tag{9.60}$$

Vector fields ξ_λ on Riemannian manifolds with the differential condition

$$\xi_{\lambda\|\eta} + \xi_{\eta\|\lambda} = 0 \tag{9.61}$$

have been discussed in Sec. 3.7. We recall that they are called Killing vector fields and indicate that a symmetry of the metric is present. In our present approximation (9.61) implies that $\hat{\tau}_\eta$ is a field of Killing vectors. We shall now show that a field of such vectors which is regular everywhere and vanishes at infinity in a space which is asymptotically pseudo-Euclidean is identically zero. By asymptotically pseudo-Euclidean, we mean here that, as the *space* coordinates x^i go to infinity, the metric becomes asymptotically pseudo-Euclidean.

Differentiation of (9.60) with respect to x^ν gives

$$\hat{\tau}_{\eta|\lambda|\nu} + \hat{\tau}_{\lambda|\eta|\nu} = 0 \tag{9.62}$$

By cyclic permutation of the indices η, λ, and ν, we then obtain the system of equations

$$\begin{aligned} \hat{\tau}_{\eta|\lambda|\nu} + \hat{\tau}_{\lambda|\nu|\eta} &= 0 \\ \hat{\tau}_{\lambda|\nu|\eta} + \hat{\tau}_{\nu|\eta|\lambda} &= 0 \\ \hat{\tau}_{\nu|\eta|\lambda} + \hat{\tau}_{\eta|\lambda|\nu} &= 0 \end{aligned} \tag{9.63}$$

These may conveniently be written in matrix form as

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\tau}_{\eta|\lambda|\nu} \\ \hat{\tau}_{\lambda|\eta|\nu} \\ \hat{\tau}_{\nu|\lambda|\eta} \end{pmatrix} = 0 \tag{9.63'}$$

since the indices of ordinary differentiation commute. This system can have a nonzero solution only if the determinant of the coefficient matrix is zero, but the determinant is clearly equal to 2; thus the only solution of the system is the null solution

$$\hat{\tau}_{\eta|\lambda|\nu} = 0 \tag{9.64}$$

By integration, we then have

$$\hat{\tau}_\eta = \text{linear function} \tag{9.65}$$

However, a linear function is either zero, constant, or infinite for large arguments. For most reasonable physical systems (such as isolated bodies or gravitational waves), the gravitational field should be asymptotically zero in at least one spatial direction, as we have already assumed. This requires that $\gamma_{\alpha\beta}$ behave similarly and that the linear function in (9.65) be identically zero; that is,

$$\hat{\tau}_\eta = 0 \tag{9.66}$$

Substituting the definition of $\hat{\tau}_\eta$ in (9.66), we have, finally,

$$\hat{\tau}_\eta = \sum_{\beta=0}^{3} (\hat{\gamma}_{\eta\beta|\beta} - \tfrac{1}{2}\hat{\gamma}_{\beta\beta|\eta}) = 0 \tag{9.67}$$

which is the set of relations on the components of $\hat{\gamma}_{\eta\lambda}$ that we desired.

We shall see in the next section that the matrices $\hat{\gamma}_{\eta\lambda}$ associated with arbitrary solutions $\gamma_{\eta\lambda}$ are indeed physically more important and meaningful than the original arbitrary solutions.

9.4 Structure of the Linearized Equations

In the preceding section we found that, with each solution of the linearized field equations $\gamma_{\eta\lambda}$, we may associate two other solutions: one is the associated Weyl solution $\gamma_{\eta\lambda}^{(w)}$, and the other is the difference between the original solution and its associated Weyl solution $\gamma_{\eta\lambda} - \gamma_{\eta\lambda}^{(w)}$, which we call $\hat{\gamma}_{\eta\lambda}$. Now we wish to show that the only solution with physical importance is the second associated solution $\hat{\gamma}_{\eta\lambda}$.

Recall from Chap. 5 that the Riemann tensor $R^{\alpha}{}_{\eta\beta\lambda}$ determines several important properties of space; in particular, a null curvature tensor is a necessary and sufficient condition that a space be Lorentzian or pseudo-Euclidean (Sec. 5.6). Therefore, as a first step in investigating the role played by the Weyl solutions in the structure of the linearized equations, we shall calculate the Riemann tensor for a metric tensor of the form $g_{\alpha\beta}^{(L)} + \epsilon\gamma_{\alpha\beta}^{(w)}$. By definition (Sec. 5.2), the Riemann tensor is

$$(9.68)\qquad R^{\alpha}{}_{\eta\beta\lambda} = \begin{Bmatrix}\alpha\\ \beta\ \eta\end{Bmatrix}_{|\lambda} - \begin{Bmatrix}\alpha\\ \eta\ \lambda\end{Bmatrix}_{|\beta} + \begin{Bmatrix}\alpha\\ \tau\ \lambda\end{Bmatrix}\begin{Bmatrix}\tau\\ \beta\ \eta\end{Bmatrix} - \begin{Bmatrix}\alpha\\ \tau\ \beta\end{Bmatrix}\begin{Bmatrix}\tau\\ \lambda\ \eta\end{Bmatrix}$$

Because of the form of the metric tensor, each Christoffel symbol is of order ϵ, so the last two terms of the Riemann tensor are of order ϵ^2 and may be deleted in the linearized theory. This leaves

$$(9.69)\qquad R^{\alpha}{}_{\eta\beta\lambda} = \begin{Bmatrix}\alpha\\ \beta\ \eta\end{Bmatrix}_{|\lambda} - \begin{Bmatrix}\alpha\\ \eta\ \lambda\end{Bmatrix}_{|\beta}$$

The Christoffel symbols have already been investigated in Sec. 9.1. To first order in ϵ we may write the Christoffel symbol as

$$(9.70)\qquad \begin{Bmatrix}\alpha\\ \beta\ \eta\end{Bmatrix} = -\frac{\epsilon}{2}(\gamma_{\alpha\beta|\eta} + \gamma_{\alpha\eta|\beta} - \gamma_{\beta\eta|\alpha})$$

Substitution of this in (9.69) gives

$$(9.71)\qquad R^{\alpha}{}_{\eta\beta\lambda} = \frac{\epsilon}{2}(\gamma_{\alpha\lambda|\eta|\beta} + \gamma_{\beta\eta|\alpha|\lambda} - \gamma_{\alpha\beta|\eta|\lambda} - \gamma_{\eta\lambda|\alpha|\beta})$$

This equation is valid for an arbitrary perturbation matrix $\gamma_{\alpha\beta}$. For the special case of a Weyl solution the matrix $\gamma_{\alpha\beta}$ has the form $\varphi_{\alpha|\beta} + \varphi_{\beta|\alpha}$.

The corresponding Riemann tensor is therefore

$$(9.72)\qquad \begin{aligned} R^{\alpha}{}_{\eta\beta\lambda} &= \frac{\epsilon}{2}(\varphi_{\alpha|\lambda|\eta|\beta} + \varphi_{\lambda|\alpha|\eta|\beta} + \varphi_{\beta|\eta|\alpha|\lambda} + \varphi_{\eta|\beta|\alpha|\lambda} \\ &\qquad - \varphi_{\alpha|\beta|\eta|\lambda} - \varphi_{\beta|\alpha|\eta|\lambda} - \varphi_{\eta|\lambda|\alpha|\beta} - \varphi_{\lambda|\eta|\alpha|\beta}) \\ &= 0 \end{aligned}$$

The Weyl solution gives rise to a null Riemann tensor (to first order in ϵ).

In the case of a more general solution $\gamma_{\alpha\beta}$, we may use the decomposition $\gamma_{\alpha\beta} = \gamma_{\alpha\beta}^{(w)} + \hat{\gamma}_{\alpha\beta}$. From (8.71) it is clear that, within our approximation, the Riemann tensor depends linearly on $\gamma_{\alpha\beta}$ and can be written as the sum of a term which depends only on $\gamma_{\alpha\beta}^{(w)}$ and a term which depends only on $\hat{\gamma}_{\alpha\beta}$; we have shown above that the term depending on $\gamma_{\alpha\beta}^{(w)}$ is identically zero, so the entire Riemann tensor depends only on $\hat{\gamma}_{\alpha\beta}$. Since the Weyl term is irrelevant in determining the curvature tensor of the Riemann space, we are led to suspect that it corresponds to a *formal* property of the linearized equations and is of no physical consequence. We shall show below that this is indeed the case; the Weyl solution stems entirely from the freedom we have in choosing a Minkowski coordinate system and can furthermore be eliminated by an appropriate choice of coordinate system.

Our assumption on the form of the metric tensor

$$(9.73)\qquad g_{\alpha\beta} = -\delta_{\alpha\beta} + \epsilon\gamma_{\alpha\beta}$$

allows a coordinate transformation of the form

$$(9.74)\qquad \bar{x}^{\mu} = x^{\mu} - \epsilon\varphi_{\mu}$$

which does not change the form of the metric (9.73). Let us show explicitly the effect of (9.74) on the metric tensor; the transformation coefficients to first order in ϵ are

$$(9.75)\qquad \frac{\partial \bar{x}^{\mu}}{\partial x^{\alpha}} = \delta^{\mu}{}_{\alpha} - \epsilon\frac{\partial \varphi_{\mu}}{\partial x^{\alpha}} \qquad \frac{\partial x^{\alpha}}{\partial \bar{x}^{\mu}} = \delta^{\alpha}{}_{\mu} + \epsilon\frac{\partial \varphi_{\alpha}}{\partial \bar{x}^{\mu}}$$

Thus, in the new system, the metric tensor is

$$(9.76)\qquad \begin{aligned} \bar{g}_{\mu\nu} &= \left(\delta^{\alpha}{}_{\mu} + \epsilon\frac{\partial\varphi_{\alpha}}{\partial\bar{x}^{\mu}}\right)\left(\delta^{\beta}{}_{\nu} + \epsilon\frac{\partial\varphi_{\beta}}{\partial\bar{x}^{\nu}}\right)(-\delta_{\alpha\beta} + \epsilon\gamma_{\alpha\beta}) \\ &= -\delta_{\mu\nu} + \epsilon\left[\gamma_{\mu\nu} - \frac{\partial\varphi_{\nu}}{\partial\bar{x}^{\mu}} - \frac{\partial\varphi_{\mu}}{\partial\bar{x}^{\nu}}\right] \\ &= -\delta_{\mu\nu} + \epsilon\,[\gamma_{\mu\nu} - \varphi_{\nu|\mu} - \varphi_{\mu|\nu}] \end{aligned}$$

Thus the metric tensor in the new system has the same form as in (9.73), but we see that a Weyl-type solution, $-(\varphi_{\nu|\mu} + \varphi_{\mu|\nu})$, has been added to $\gamma_{\alpha\beta}$. Hence the influence of the arbitrariness of the coordinate system due to the freedom of choosing the function φ_μ in (9.74) is to add a Weyl-type solution to $\gamma_{\alpha\beta}$.

By a proper choice of the functions φ_μ used in the above discussion, it is clear that the Weyl part of any solution $\gamma_{\alpha\beta} = \gamma_{\alpha\beta}^{(w)} + \hat{\gamma}_{\alpha\beta}$ can be eliminated by the simple coordinate transformation (9.74). Moreover, if $\gamma_{\alpha\beta}$ were entirely a Weyl type of solution, we could easily go to a system where $\gamma_{\alpha\beta} = -\delta_{\alpha\beta}$, which would imply a pseudo-Euclidean space. This agrees with the fact that the Riemann tensor corresponding to a Weyl solution is identically zero, for, as we discussed in Chap. 5, a space with a null curvature tensor is pseudo-Euclidean, and vice versa.

Let us consider for a moment our results and their significance in the structure of the linearized equations. We found that the presence of the Weyl solution in the metric tensor $g_{\alpha\beta}^{(L)} + \gamma_{\alpha\beta}^{(w)}$ gave rise to a null Riemann tensor. Therefore the Weyl solution cannot correspond to a gravitational field. Indeed, we have seen that the transformation (9.73) can explicitly remove the Weyl solution from the metric tensor. We must therefore consider the Weyl solution to be a formal property of the coordinate system we are using, and not a physical property of space. A problem, therefore, confronts us: How can we separate out the Weyl solutions so that, in general, we deal only with physically meaningful solutions? One method of consistently discarding the Weyl solutions would be to obtain first a solution $\gamma_{\alpha\beta}$, then compute its associated Weyl solution $\gamma_{\alpha\beta}^{(w)}$ using (9.53) and (9.54), and then subtract $\gamma_{\alpha\beta}^{(w)}$ from $\gamma_{\alpha\beta}$ to form $\hat{\gamma}_{\alpha\beta}$, which is the physically meaningful solution. However, there is a much simpler course open to us whereby the Weyl solutions are automatically excluded from consideration and never appear in our calculations at all. We found in the preceding sections that the solution $\hat{\gamma}_{\alpha\beta}$ alone obeys several sets of equations—the second-order D'Alembertian differential equation

$$\Box^2\hat{\gamma}_{\alpha\beta} = 0 \tag{9.77}$$

and a set of first-order differential equations relating the components

$$\sum_{\beta=0}^{3} (\hat{\gamma}_{\eta\beta|\beta} - \tfrac{1}{2}\hat{\gamma}_{\beta\beta|\eta}) = c_\eta \tag{9.78}$$

in which the constant c_η is zero for Euclidean boundary conditions at infinity (as we discussed in Sec. 9.3) and depends on the specific boundary conditions otherwise. These two sets of equations form a very elegant restatement of the linearized equations since they are indeed equivalent to the linearized equations with the Weyl solution automatically discarded.

Let us note, finally, that one of the most important results of the linearized theory is the fact that the physically interesting solutions $\hat{\gamma}_{\alpha\beta}$ of the linearized equations satisfy the wave equation (9.77). Hence perturbations in the metric field (gravitational "waves") satisfy the same differential equation as electromagnetic phenomena. However, the 10 components of the metric field $\hat{\gamma}_{\alpha\beta}$ are also coupled by the four relations (9.78), which implies that properties such as the polarization configurations of the metric field may differ considerably from the electromagnetic field. The fact that electromagnetic and gravitational disturbances follow the same paths (null geodesics) was shown in general, in Chap. 7, with no approximations. However, the general approach did not reveal the four relations (9.78) which couple the components of the metric field. Thus both the general approach and the linearized approach to the theory of gravitational waves have certain advantages and certain drawbacks; the use of both together in the study of gravitational propagation is clearly of value.

9.5 Gravitational Waves

A great deal of thought has been devoted in recent years to the theoretical analysis of gravitational waves, but the detection of such waves is a very difficult experimental task only recently attempted. We shall say more about these efforts at detection later in this section. For clarity our brief theoretical analysis of gravitational waves will deal only with plane-wave solutions to the linearized field equations.

Let us begin by asking what theoretical motivation there is to believe that gravitational waves exist. The linearized equations may be written in the concise form

$$\Box^2\gamma_{\mu\nu} = 0 \tag{9.79a}$$

$$\sum_{\beta=0}^{3} (\gamma_{\eta\beta|\beta} - \tfrac{1}{2}\gamma_{\beta\beta|\eta}) = 0 \tag{9.79b}$$

as we showed in the preceding section. We have now dropped the $\hat{\gamma}_{\mu\nu}$ notation for convenience, and we shall assume Euclidean boundary conditions at infinity. One might be tempted to interpret (9.79*a*) as indicating that gravitational effects propagate as waves with velocity c. This, however, is open to the objection that the perturbation $\gamma_{\mu\nu}$ is linked to an arbitrary coordinate system, and therefore the existence of a nonzero metric perturbation is not an invariant indication of the existence of a gravitational field. A better argument can be made if we note that the Riemann tensor has the first-order form

$$R^{\alpha}{}_{\eta\beta\lambda} = \frac{\epsilon}{2}(\gamma_{\alpha\lambda|\eta|\beta} + \gamma_{\beta\eta|\alpha|\lambda} - \gamma_{\alpha\beta|\eta|\lambda} - \gamma_{\eta\lambda|\alpha|\beta}) \tag{9.80}$$

as we found in Eq. (9.71). We know that the part of this tensor corresponding to the extraneous Weyl solution is identically zero, so by using Eq. (9.79*a*), we see that

$$\Box^2 R^{\alpha}{}_{\eta\beta\lambda} = 0 \tag{9.81}$$

Thus the Riemann tensor, which gives an absolute criterion for the existence of a gravitational field, itself obeys the wave equation. It follows that, in the linearized theory, gravitational effects propagate with velocity c.

It should be noted carefully that the results of the above paragraph do not indicate whether or not gravitational radiation, which involves an energy transfer, exists. We may conclude only that the effects of gravity propagate at velocity c via the wave equation.

Let us now investigate the general properties of a plane-wave solution of Eqs. (9.79). It will be convenient, first, to study the general transformation properties of (9.79); specifically, we wish to obtain the conditions under which Eqs. (9.79) are invariant. For a general first-order coordinate transformation

$$\bar{x}^{\alpha} = x^{\alpha} - \epsilon\varphi_{\alpha}(x) \tag{9.82}$$

we found in the preceding section that, to first order in ϵ,

$$\frac{\partial x^{\alpha}}{\partial \bar{x}^{\mu}} = \delta^{\alpha}{}_{\mu} + \epsilon\varphi_{\alpha|\mu} \tag{9.83}$$

[Eq. (9.75)]. An easy calculation then gives

$$\bar{g}_{\mu\nu} = g_{\mu\nu} - \epsilon(\varphi_{\mu|\nu} + \varphi_{\nu|\mu}) \tag{9.84}$$

or

$$\bar{\gamma}_{\mu\nu} = \gamma_{\mu\nu} - (\varphi_{\mu|\nu} + \varphi_{\nu|\mu}) \tag{9.85}$$

From (9.83) it is evident that $\partial/\partial\bar{x}^{\lambda} = \partial/\partial x^{\lambda} + O(\epsilon)$. Thus the D'Alembertian operator $\Box^2$ is an invariant to zeroth order, and we can write (9.79*a*) in the barred system as

$$\bar{\Box}^2\bar{\gamma}_{\mu\nu} = \Box^2\gamma_{\mu\nu} - \Box^2(\varphi_{\mu|\nu} + \varphi_{\nu|\mu}) \tag{9.86}$$

Similarly, (9.79*b*) becomes

$$\sum_{\beta=0}^{3}(\bar{\gamma}_{\eta\beta|\beta} - \tfrac{1}{2}\bar{\gamma}_{\beta\beta|\eta}) = \sum_{\beta=0}^{3}(\gamma_{\eta\beta|\beta} - \tfrac{1}{2}\gamma_{\beta\beta|\eta}) + \Box^2\varphi_{\eta} \tag{9.87}$$

Thus it is clear that the linearized equations (9.79) will be invariant under the transformation (9.82) if and only if $\Box^2\varphi_{\eta} = 0$. This fact is of interest by itself and will also be useful later in this section.

In the linearized theory, plane-wave solutions have the interesting property that the Riemann tensor is highly degenerate; by this we mean that not all the components of the metric tensor occur in $R^{\alpha}{}_{\eta\beta\lambda}$. We shall demonstrate this for a plane wave in the x direction. Such a wave is characterized by the fact that all variables $\gamma_{\alpha\beta}$ depend only on the coordinates $x^0 = ict$ and $x^1 = x$, so that the metric-tensor components have a vanishing y and z derivative. That is, $\gamma_{\mu\nu|2} = \gamma_{\mu\nu|3} = 0$. We use this condition and write down the 21 independent components of $R_{\alpha\eta\beta\lambda}$. Observe that, within our approximation, $R_{\alpha\eta\beta\lambda} = -R^{\alpha}{}_{\eta\beta\lambda}$. Since

$$R^{\alpha}{}_{\eta\beta\lambda} = \frac{\epsilon}{2}(\gamma_{\alpha\lambda|\eta|\beta} + \gamma_{\beta\eta|\alpha|\lambda} - \gamma_{\alpha\beta|\eta|\lambda} - \gamma_{\eta\lambda|\alpha|\beta}) \tag{9.88}$$

we obtain the following three groups of terms:

$$R^1{}_{223} = R^1{}_{323} = R^1{}_{023} = R^2{}_{323} = R^2{}_{320} = R^2{}_{330} = 0 \tag{9.89a}$$

$$\left\{\begin{aligned} R^1{}_{030} &= \frac{\epsilon}{2}(\gamma_{30|1|0} - \gamma_{13|0|0}) \qquad & R^1{}_{020} &= \frac{\epsilon}{2}(\gamma_{20|1|0} - \gamma_{12|0|0}) \\ R^1{}_{310} &= \frac{\epsilon}{2}(\gamma_{31|1|0} - \gamma_{30|1|1}) \qquad & R^1{}_{210} &= \frac{\epsilon}{2}(\gamma_{21|1|0} - \gamma_{20|1|1}) \\ R^1{}_{010} &= \epsilon(\gamma_{10|0|1} - \tfrac{1}{2}\gamma_{11|0|0} - \tfrac{1}{2}\gamma_{00|1|1}) \end{aligned}\right. \tag{9.89b}$$

$$\left\{\begin{aligned} R^2{}_{020} &= -\frac{\epsilon}{2}\gamma_{22|0|0} \qquad & R^2{}_{030} &= -\frac{\epsilon}{2}\gamma_{23|0|0} \qquad & R^3{}_{030} &= -\frac{\epsilon}{2}\gamma_{33|0|0} \\ R^1{}_{220} &= \frac{\epsilon}{2}\gamma_{22|0|1} \qquad & R^1{}_{230} &= \frac{\epsilon}{2}\gamma_{23|0|1} \qquad & R^1{}_{320} &= \frac{\epsilon}{2}\gamma_{32|0|1} \\ R^1{}_{330} &= \frac{\epsilon}{2}\gamma_{33|0|1} \qquad & R^1{}_{212} &= -\frac{\epsilon}{2}\gamma_{22|1|1} \qquad & R^1{}_{213} &= -\frac{\epsilon}{2}\gamma_{23|1|1} \\ R^1{}_{313} &= -\frac{\epsilon}{2}\gamma_{33|1|1} \end{aligned}\right. \tag{9.89c}$$

The values of all components $R^\alpha{}_{\eta\beta\lambda}$ can be obtained from these formulas by the above-mentioned approximate identity $R_{\alpha\eta\beta\lambda} = -R^\alpha{}_{\eta\beta\lambda}$ and the symmetries of the Riemann tensor. If we now impose the linearized field equations in the form $R_{\eta\lambda} = 0$, we obtain, for instance,

$$R_{13} = R^\alpha{}_{1\alpha 3} = R^0{}_{103} = 0 \tag{9.90}$$

A similar consideration of the remaining field equations yields the result that each of the components of the curvature tensor in group (9.89*b*) vanishes identically; thus only the components in group (9.89*c*) are nonzero. These components involve only γ_{22}, γ_{23}, γ_{32}, and γ_{33}; thus the curvature tensor is a function of only these elements of the metric tensor.

The formal result of the foregoing paragraph has a very interesting consequence. If we write the perturbation $\gamma_{\mu\nu}$ in two parts,

$$\gamma_{\mu\nu} = \gamma_{\mu\nu}(1) + \gamma_{\mu\nu}(2)$$

$$\gamma_{\mu\nu}(1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{22} & \gamma_{23} \\ 0 & 0 & \gamma_{32} & \gamma_{33} \end{pmatrix} \qquad \gamma_{\mu\nu}(2) = \begin{pmatrix} \gamma_{00} & \gamma_{01} & \gamma_{02} & \gamma_{03} \\ \gamma_{10} & \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{20} & \gamma_{21} & 0 & 0 \\ \gamma_{30} & \gamma_{31} & 0 & 0 \end{pmatrix} \tag{9.91}$$

the curvature tensor of $\gamma_{\mu\nu}(2)$ will be identically zero. This indicates that there should exist a coordinate system where $\gamma_{\mu\nu}$ has only γ_{22}, γ_{23}, γ_{32}, and γ_{33} as nonzero components; that is, $\gamma_{\mu\nu}$ is a pure $\gamma_{\mu\nu}(1)$-type solution. Indeed such a coordinate system will be constructed in the following paragraphs. We shall refer to such a form as a canonical-wave solution.

To verify the above assertion we shall now explicitly solve (9.79) for a plane wave in the x^1 direction and put the result in canonical form by a coordinate transformation. Equation (9.79*a*) is automatically satisfied if we choose the plane-wave functional dependence

$$\gamma_{\mu\nu} = \gamma_{\mu\nu}(x^1 - ct) = \gamma_{\mu\nu}(x^1 + ix^0)$$

Now denote $\sum_{\beta=0}^{3} \gamma_{\beta\beta} = \operatorname{Tr}\gamma$ by Γ, so the four components of (9.79*b*) may be written

$$\begin{aligned} \gamma_{00|0} + \gamma_{01|1} - \tfrac{1}{2}\Gamma_{|0} &= 0 \\ \gamma_{10|0} + \gamma_{11|1} - \tfrac{1}{2}\Gamma_{|1} &= 0 \\ \gamma_{20|0} + \gamma_{21|1} &= 0 \\ \gamma_{30|0} + \gamma_{31|1} &= 0 \end{aligned} \tag{9.92}$$

Because of the functional form of $\gamma_{\mu\nu}$, we can simplify these equations by noting that

$$\begin{aligned} \gamma_{\mu\nu|1} &= \gamma'_{\mu\nu} \\ \gamma_{\mu\nu|0} &= i\gamma'_{\mu\nu} \end{aligned} \tag{9.93}$$

where the prime denotes differentiation with respect to the argument $x^1 + ix^0$. Then we obtain for (9.92) the form

$$\begin{aligned} i\gamma'_{00} + \gamma'_{01} - \frac{i}{2}\Gamma' &= 0 \\ i\gamma'_{10} + \gamma'_{11} - \tfrac{1}{2}\Gamma' &= 0 \\ i\gamma'_{20} + \gamma'_{21} &= 0 \\ i\gamma'_{30} + \gamma'_{31} &= 0 \end{aligned} \tag{9.94}$$

Furthermore, since the $\gamma_{\mu\nu}$ all vanish at infinity for Euclidean boundary conditions, the entire set can be integrated at once merely by dropping the primes. Then multiplying the second equation by i and adding the first, we obtain

$$\gamma_{22} = -\gamma_{33} \qquad \Gamma = \operatorname{Tr}\gamma = \gamma_{00} + \gamma_{11} \tag{9.95}$$

Either the first or second equation in (9.94) may then be solved for γ_{01}:

$$\gamma_{01} = \frac{i}{2}(\gamma_{11} - \gamma_{00}) \tag{9.96}$$

Finally, the last two equations give

$$\gamma_{20} = i\gamma_{21} \qquad \gamma_{30} = i\gamma_{31} \tag{9.97}$$

Collecting the results (9.95) to (9.97), we write $\gamma_{\mu\nu}$ as

$$\gamma_{\mu\nu} = \left(\begin{array}{cc|cc} \gamma_{00} & \frac{i}{2}(\gamma_{11} - \gamma_{00}) & i\gamma_{12} & i\gamma_{13} \\ \frac{i}{2}(\gamma_{11} - \gamma_{00}) & \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \hline i\gamma_{12} & \gamma_{12} & \gamma_{22} & \gamma_{23} \\ i\gamma_{13} & \gamma_{13} & \gamma_{23} & -\gamma_{22} \end{array}\right) \tag{9.98}$$

This is the most general solution in an arbitrary nearly Lorentzian coordinate system.

We wish now to put (9.98) in canonical form by going to a coordinate system where the extraneous components of (9.98) vanish, i.e., where $\gamma_{\mu\nu}$ has only a (2,3) subblock as in (9.91). In addition, we shall restrict ourselves to transformations of the form (9.82), with $\Box^2\varphi_\alpha = 0$, so that the linearized equations (9.79) retain their form in the barred system. If we demand that $\bar\gamma_{00} = \bar\gamma_{11} = \bar\gamma_{12} = \bar\gamma_{13} = 0$, then the transformation equation (9.85) implies that the φ_μ must obey the first-order equations

$$\varphi_{0|0} = \tfrac{1}{2}\gamma_{00} \tag{9.99a}$$

$$\varphi_{1|1} = \tfrac{1}{2}\gamma_{11} \tag{9.99b}$$

$$\varphi_{1|2} + \varphi_{2|1} = \gamma_{12} \tag{9.99c}$$

$$\varphi_{1|3} + \varphi_{3|1} = \gamma_{13} \tag{9.99d}$$

If we now choose the φ_μ to have the functional form $\varphi_\mu(x^1 + ix^0)$, the subsidiary condition $\Box^2\varphi_\mu = 0$ is clearly satisfied. In order to satisfy (9.99a), we need merely choose a function $F(Z)$ such that $F'(Z) = \gamma_{00}(Z)$; then $\varphi_0(x^1 + ix^0) = (-i/2)F(x^1 + ix^0)$ will satisfy (9.99a). Indeed, if we choose functions $F(Z)$, $G(Z)$, $H(Z)$, and $K(Z)$ such that

$$\begin{aligned} F'(Z) &= \gamma_{00}(Z) \qquad & G'(Z) &= \gamma_{11}(Z) \\ H'(Z) &= \gamma_{12}(Z) \qquad & K'(Z) &= \gamma_{13}(Z) \end{aligned} \tag{9.100}$$

then it is evident that Eqs. (9.99) and the subsidiary condition are satisfied by the following functions:

$$\begin{aligned} \varphi_0 &= \frac{-i}{2}F(x^1 + ix^0) \qquad & \varphi_1 &= \tfrac{1}{2}G(x^1 + ix^0) \\ \varphi_2 &= H(x^1 + ix^0) \qquad & \varphi_3 &= K(x^1 + ix^0) \end{aligned} \tag{9.101}$$

Note, furthermore, that the choice (9.101) leaves the (2,3) subblock unchanged because of the functional form of the φ_μ. This completes the demonstration that an allowed coordinate transformation always exists which will put $\gamma_{\mu\nu}$ into the canonical form

$$\gamma_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{22} & \gamma_{23} \\ 0 & 0 & \gamma_{23} & -\gamma_{22} \end{pmatrix} \tag{9.102}$$

To close this section we shall briefly study the motion of a test particle which moves in the metric field (9.102) of the plane wave. It describes

the geodesic motion

$$\ddot{x}^\mu + \begin{Bmatrix} \mu \\ \alpha\ \beta \end{Bmatrix} \dot{x}^\alpha \dot{x}^\beta = 0 \tag{9.103}$$

Because of the particularly simple structure of the metric (9.102), we can easily obtain the Christoffel symbols. We find that, to order ϵ,

$$\begin{Bmatrix} 0 \\ \alpha\ \beta \end{Bmatrix} = -\frac{\epsilon}{2}(\gamma_{0\alpha|\beta} + \gamma_{0\beta|\alpha} - \gamma_{\alpha\beta|0}) = \frac{\epsilon}{2}\gamma_{\alpha\beta|0} \tag{9.104}$$

since all $\gamma_{0\alpha}$ are zero. Similarly,

$$\begin{Bmatrix} 1 \\ \alpha\ \beta \end{Bmatrix} = \frac{\epsilon}{2}\gamma_{\alpha\beta|1} \tag{9.105}$$

On the other hand, we know that $\gamma_{\alpha\beta}$ is a function of $x^1 + ix^0$, so that $\gamma_{\alpha\beta|0} = i\gamma_{\alpha\beta|1}$, from which we obtain

$$i\begin{Bmatrix} 1 \\ \alpha\ \beta \end{Bmatrix} = \begin{Bmatrix} 0 \\ \alpha\ \beta \end{Bmatrix} \tag{9.106}$$

From (9.103) and (9.106) it immediately follows that

$$i\ddot{x}^0 + \ddot{x}^1 = 0 \tag{9.107}$$

which has a first integral which will be useful:

$$i\dot{x}^0 + \dot{x}^1 = A \tag{9.108}$$

Let us next display the equation of motion (9.103) for $\mu = 1$:

$$\ddot{x}^1 + \frac{\epsilon}{2}[\gamma'_{22}((\dot{x}^2)^2 - (\dot{x}^3)^2) + 2\gamma'_{23}\dot{x}^2\dot{x}^3] = 0 \tag{9.109}$$

We also desire the equations of motion for $\mu = 2$ and 3. Since

$$\begin{Bmatrix} k \\ \alpha\ \beta \end{Bmatrix} = -\frac{\epsilon}{2}(\gamma_{\alpha k|\beta} + \gamma_{\beta k|\alpha}) \qquad k = 2, 3 \tag{9.110}$$

we obtain

$$\ddot{x}^k - \epsilon\gamma_{\alpha k|\beta}\dot{x}^\alpha\dot{x}^\beta = 0 \qquad k = 2, 3 \tag{9.111}$$

in which α runs only over 2 and 3, and β runs only over 0 and 1. Since $\gamma_{\alpha k|0} = i\gamma_{\alpha k|1} = i\gamma'_{\alpha k}$, we find

$$\ddot{x}^k = \epsilon\gamma'_{\alpha k}(i\dot{x}^0 + \dot{x}^1)\dot{x}^\alpha = \epsilon(i\dot{x}^0 + \dot{x}^1)(\gamma'_{2k}\dot{x}^2 + \gamma'_{3k}\dot{x}^3) \tag{9.112}$$

By virtue of (9.108), we therefore find

$$\ddot{x}^2 = \epsilon A(\gamma'_{22}\dot{x}^2 + \gamma'_{23}\dot{x}^3) \tag{9.113a}$$

$$\ddot{x}^3 = \epsilon A(\gamma'_{23}\dot{x}^2 + \gamma'_{33}\dot{x}^3) = \epsilon A(\gamma'_{23}\dot{x}^2 - \gamma'_{22}\dot{x}^3) \tag{9.113b}$$

The equations of motion (9.108), (9.109), and (9.113) allow us to analyze the nature of gravitational waves in a very enlightening way. We shall assume, as always, that the gravitational fields we deal with are weak and furthermore that the velocities of all particles are small compared to c. Then we have approximately

$$\dot{x}^0 = ic\frac{dt}{ds} \cong i \qquad \dot{x}^i \cong \frac{v^i}{c} \tag{9.114}$$

Equation (9.108) then reads

$$-1 + \frac{v^1}{c} \cong -A \tag{9.115}$$

so that v^1 must remain a constant. This is consistent with (9.109), which becomes

$$\ddot{x}^1 = -\frac{\epsilon}{2}\left[\gamma'_{22}\left(\left(\frac{v^2}{c}\right)^2 - \left(\frac{v^3}{c}\right)^2\right) + 2\gamma'_{23}\frac{v^2v^3}{c^2}\right] = O\left(\frac{\epsilon v^2}{c^2}\right) \tag{9.116}$$

so that $\ddot{x}^1$ is indeed quite small. Equations (9.113) become

$$\begin{aligned} \ddot{x}^2 &= \epsilon A(\gamma'_{22}\dot{x}^2 + \gamma'_{23}\dot{x}^3) = O\left(\frac{\epsilon v}{c}\right) \\ \ddot{x}^3 &= \epsilon A(\gamma'_{23}\dot{x}^2 + \gamma'_{33}\dot{x}^3) = O\left(\frac{\epsilon v}{c}\right) \end{aligned} \tag{9.117}$$

Thus to lowest order in the velocity there is no acceleration of the particle at all!

This conclusion does not, however, mean that particles are not *moved* by the gravitational wave, only that their *coordinates* do not change. In fact there is a definite physical displacement of two particles relative to

each other as a gravitational wave passes. To see this recall that the physical distance between two bodies whose coordinate separation is dx^i along the i direction is given by $\sqrt{g_{ii}}\,dx^i$, as we discussed in Sec. 4.2. Consider first the case of a wave for which $\gamma_{23} = 0$; we have a line element given by

$$ds^2 = c^2\,dt^2 - dx^2 - (1 - \epsilon\gamma_{22})\,dy^2 - (1 + \epsilon\gamma_{22})\,dz^2 \tag{9.118}$$

Thus the physical separations between test particles at rest in the coordinate system will change as the wave passes. For example, when $\gamma_{22} > 0$, the y separation decreases and the z separation increases, as shown in Fig. 9.1 for a group of four free test particles.

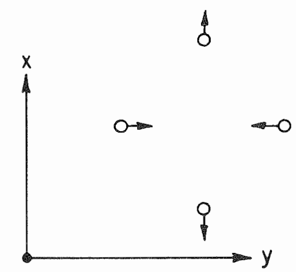


Fig. 9.1
A gravitational wave with $\gamma_{22} \neq 0$ and $\gamma_{23} = 0$ produces relative physical displacements in a group of particles as shown for the case of four test particles. A wave with $\gamma_{23} \neq 0$ and $\gamma_{22} = 0$ produces the same effect but with the axes rotated 45°.

A wave for which $\gamma_{22} = 0$ has a very beautiful relation to the above case for which $\gamma_{23} = 0$. The line element is

$$ds^2 = c^2\,dt^2 - dx^2 - dy^2 - dz^2 - 2\epsilon\gamma_{23}\,dy\,dz \tag{9.119}$$

We now perform a local coordinate rotation through an angle $\pi/4$ in the yz plane to get new coordinate intervals

$$d\bar{y} = \frac{1}{\sqrt{2}}(dy - dz) \qquad d\bar{z} = \frac{1}{\sqrt{2}}(dy + dz) \tag{9.120}$$

and a new line element

$$ds^2 = c^2\,dt^2 - dx^2 - (1 - \epsilon\gamma_{23})\,d\bar{y}^2 - (1 + \epsilon\gamma_{23})\,d\bar{z}^2 \tag{9.121}$$

Since this is precisely the same form as (9.118), we see that this second type of wave produces the same kind of effect as the first but with the axes rotated by 45°. Clearly any wave field is a superposition of these two types of waves.

The above situation is reminiscent of a similar situation in electrodynamics. A plane electromagnetic wave has two possible polarization directions, corresponding to two independent axes perpendicular to each other and to the direction of propagation of the wave. A charged particle at rest is accelerated by such a wave predominantly in the direction of its polarization. Thus the two polarization states of the electromagnetic wave produce accelerations that have the same magnitude but lie at right angles to each other. In the gravitational case we have a very similar situation, except that we must look not at the acceleration of a single particle but at the separation of two particles and the independent polarization states lie at 45° to each other.

It is important to note the fundamentally different nature of the motion of test bodies in the time-independent approximate metric of Sec. 4.3, which has nonzero γ_{00}, and the wave-type metric we have just studied, which has $\gamma_{00} = 0$. In the former case, motion resembles classical motion in a Newtonian field, but in classical gravitational theory there is no analogue of the field and the motion we have just studied: gravitational waves represent a qualitative difference between classical and relativistic gravitational theory. The difference between the metric fields is analogous to the difference between Coulomb and radiation fields in electromagnetic theory.

The detection of gravitational-wave pulses has been reported by Weber (1970a, 1970b). His detection apparatus consists of several rigid aluminum bars, each of which undergoes stresses and strains as a wave pulse passes over it. These bars are located many miles apart, and the signals are put in coincidence to reduce the contribution of noise. Although the displacements involved are very small, of order 10^{-14} cm, they are indirectly detectable through piezoelectric devices placed around the bar. A bar detector responds differently to waves coming from different directions and thus acts as a directional antenna. Early data of Weber suggested that the source of the pulses lay in the direction of the galactic center. Other workers have obtained negative results, detecting no pulses, apparently contradicting Weber's results (Levine and Garwin, 1973; Tyson, 1973). If future experiments were to confirm Weber's results, a difficult theoretical problem would arise in accounting for the origin of such a large number of intense pulses (Field et al., 1969). Some interesting sources of gravitational waves which have been studied theoretically are bodies in high-velocity orbits around black holes, rotating stars undergoing gravitational collapse, and supernova explosions, none of which appear to be capable of producing enough energy to be detected by any of the presently operating antennas. However, with a sufficient

increase in the sensitivity of gravitational-wave antennas we may expect the opening of a new field, gravitational astronomy.

Exercises

9.1 In the linearized theory solutions of the field equations may be superposed. That is, a linear combination of solutions is itself a solution. Superpose solutions of the form obtained in Sec. 9.2 to obtain new solutions for the following cases: (*a*) two point particles lying on the x axis, one at $x = d$ and the other at $x = -d$; (*b*) a continuous line of mass density $\rho(x)$ g/cm, extending from $x = -d$ to $x = d$.

9.2 Show that in the linearized theory any localized distribution of matter, e.g., contained within a sphere of radius d, has a metric which tends to (9.44) for $r \gg d$.

9.3 Obtain explicit solutions for the gravitational wave equations (9.79*a*) in which the metric components are of the form $\cos(\omega t - \mathbf{k} \cdot \mathbf{x})$ or $\sin(\omega t - \mathbf{k} \cdot \mathbf{x})$. Show that ω may be interpreted as the radian frequency of the waves and $\mathbf{k}$ as the propagation direction. What is the magnitude of $\mathbf{k}$? The subsidiary condition (9.79*b*) is guaranteed by the canonical form (9.102). Show that the general solution of the system (9.79) may be written as a superposition of such plane waves, integrated over frequencies, and summed over the two polarization states.

9.4 In Chap. 5 we related the components $R^i{}_{0j0}$ of the Riemann tensor to classical tidal forces, or derivatives of the Newtonian force. Use this correspondence to determine the force exerted on a rigid body by a gravitational wave.

9.5 Discuss how a simple system of masses and springs, e.g., two masses connected by a spring, would respond to a gravitational wave. Study the motion for a single sinusoidal wave as discussed in Exercise 9.3, in particular the response as a function of the wave frequency.

9.6 Apply the equation of geodesic deviation to the plane-wave metric to obtain the quadrupole deformations produced on a system of test particles by the two polarization states.

Problems

9.1 We have studied gravitational waves in vacuum without regard to their sources. This problem is discussed by Weber (1961) in detail (see also Sec. 10.5). Show that a system of moving masses will, in general, radiate gravitational waves, with a source strength proportional to the second time derivative of the quadrupole moment.

9.2 For a system with a sinusoidally varying quadrupole moment show that the power radiated is of order $P \sim \kappa^2 I^2 \omega^6/c^5$, where I is the moment of inertia and ω is the radian frequency. (An expression for the energy density of the gravitational field may be obtained from Sec. 10.2 or from Weber's book.)

9.3 Discuss the gravitational radiation produced when a rotating star of roughly solar mass collapses to a black hole (see Chap. 14). Use order-of-magnitude estimates.

Bibliography

Ehlers, J., and W. Kundt (1962): Exact Solutions of the Gravitational Field Equations, in L. Witten (ed.), "Gravitation: An Introduction to Current Research," New York, pp. 49–101.

Einstein, A. (1914): Die formalen Grundlagen der allgemeinen Relativitätstheorie, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 1030–1085.

Einstein, A. (1915a): Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 831–839.

Einstein, A. (1915b): Die Feldgleichungen der Gravitation, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 844–847.

Einstein, A. (1916): Näherungsweise Integration der Feldgleichungen der Gravitation, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 688–696.

Einstein, A. (1918): Über Gravitationswellen, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 154–167.

Field, G. B., M. J. Rees, and D. W. Sciama (1969): Astronomical Significance of Mass Loss by Gravitational Radiation, *Comments Astrophys. Space Phys.*, **1**:187.

Landau, L., and E. M. Lifshitz (1962): "The Classical Theory of Fields," 2d ed., Reading, Mass.

Levine, J. L., and R. L. Garwin (1973a): Absence of Gravity Wave Signals in a Bar at 1695 Hz, *Phys. Rev. Letters*, **31**:173.

Levine, J. L., and R. L. Garwin (1973b): Single Gravity Wave Detector Results Contrasted with Previous Coincidence Detection, *Phys. Rev. Letters*, **31**:176.

Pirani, F. A. E. (1962): Gravitational Radiation, in L. Witten (ed.), "Gravitation: An Introduction to Current Research," New York, pp. 199–226.

Tyson, J. A. (1973): Null Search for Bursts of Gravitational Radiation, *Phys. Rev. Letters*, **31**:326.

Weber, J. (1961): "General Relativity and Gravitational Waves," New York.

Weber, J. (1964): Gravitational Waves, in B. de Witt (ed.), "General Relativity, Groups and Topology," New York.

Weber, J. (1969): Evidence for Discovery of Gravitational Radiation, *Phys. Rev. Letters*, **22**:1320.

Weber, J. (1970a): Gravitational Radiation Experiments, *Phys. Rev. Letters*, **24**:276.

Weber, J. (1970b): Anisotropy and Polarization in the Gravitational Radiation Experiments, *Phys. Rev. Letters*, **25**:180.

Weinberg, S. (1972): "Gravitation and Cosmology," New York, sec. 10.7 on detection of gravitational radiation.

Weyl, H. (1950): "Space, Time, Matter," New York (1st ed., 1918).

CHAPTER

10

The Gravitational Field Equations for Nonempty Space

In Chap. 5 we presented the gravitational field equations for free space,

$$R_{\gamma\eta} = 0 \tag{10.1}$$

which were first proposed by Einstein. Then, in Chap. 9, we found that these equations reduce in the limit of weak fields to Laplace's equation for the classical gravitational potential φ,

$$\sum_{i=1}^{3} \varphi_{|i|i} = 0 \tag{10.2}$$

and may therefore be considered to be a generalization of the classical theory for free space. In this chapter we shall make a similar generalization of the classical gravitational equation for nonempty space, Poisson's equation:

$$\sum_{i=1}^{3} \varphi_{|i|i} = 4\pi\rho\kappa \qquad \kappa = 6.67 \times 10^{-8}\ \text{dyne-cm}^2/\text{g}^2 \tag{10.3}$$

The scalar ρ denotes the density of matter in space, and κ is the gravitational constant. Observe that a distribution of matter with density $\rho(x)$ in Euclidean space has the gravitational potential

$$\varphi(x) = -\kappa \int \frac{\rho(x')\,d^3x'}{|x - x'|} \tag{10.4}$$

and that (10.3) is a consequence of the identity

$$\nabla^2 \int \frac{\rho(x')}{|x - x'|}\, d^3x' = -4\pi\rho(x) \tag{10.5}$$

We shall find relativistic field equations for space containing matter (or energy) which have a tensor form analogous to (10.1) and reduce to (10.1) in the case of empty space. Using the relativistic field equations, we shall then investigate the classical limit for weak fields and show that (10.3) is a first approximation.

10.1 The Energy-Momentum Tensor

The classical equation (10.3) relates the behavior of the potential function φ to the density of *matter* in space ρ. In relativity theory, however, we cannot simply speak of the density of matter in space; we must also include the energy density, since, as Einstein has shown, matter and energy are indistinguishable with regard to their inertial properties: That is, $E = mc^2$. In dealing with nonempty space we shall therefore lump together the matter, radiant energy, elastic energy, etc., and speak of the *energy content* of space. This term, however, does *not* include gravitational energy.

We shall express the influence of matter and field energy in the form of a tensor $T^{\mu\nu}$ which is called the energy-momentum tensor. In this book we shall give $T^{\mu\nu}$ the dimensions of a pure mass density, grams per cubic centimeter. One could equally well give it the dimensions of energy density merely by multiplying the numerical values by c^2.

Following the development of the free-space field equations in Chap. 5 we shall seek a second-rank tensor equation of the general form

$$\begin{pmatrix}\text{Tensor representing}\\ \text{geometry of space}\end{pmatrix} = \begin{pmatrix}\text{tensor representing}\\ \text{energy content of space}\end{pmatrix} \tag{10.6}$$

These field equations must satisfy two limit requirements: they must be equivalent to *Poisson's equation* (10.3) in the *limit of weak fields* and must reduce to *Einstein's free-space field equations* (10.1) when the *energy density in space is zero*. We shall begin the development by considering in this section a class of tensors which describe the energy content of space and are suitable for use in the right side of the symbolic field equation (10.6). Since the free-space field equations (10.1) involve tensors of second rank, it is natural to limit our discussion to symmetric second-rank tensors.

Furthermore, we shall begin our work in a *flat* Riemann space and use the ideas of special relativity in order to retain simplicity until the end of this section. In order to develop machinery to describe matter in tensor form, we begin by recasting well-known equations of classical physics in tensor form and then look for common and characteristic features in the equations.

Let us first consider the simplest kind of energy field: a field of noninteracting incoherent matter. Such a field may be characterized by a scalar proper-density field $\rho_0(x)$ and a four-vector field of flow $u^\mu(x)$. Recall that the proper density is the density which would be measured by an observer moving *with* the flow. The four-velocity flow $u^\mu(x)$ is to be interpreted as follows: The element of matter which occupies the point x^μ of space-time has a motion $x^\mu(s)$ such that $dx^\mu/ds = u^\mu(x)$. Using these two characteristics of the matter field, the simplest second-rank tensor field we can construct is

$$T^{\mu\nu} = \rho_0(x)u^\mu(x)u^\nu(x) \tag{10.7}$$

We shall call this the energy-momentum tensor of the matter field or, more concisely, the matter tensor.

In order to obtain a physical interpretation of this tensor field, let us use the familiar coordinates of special relativity (ct,x,y,z) and the usual Lorentz metric tensor. The component T^{00} of the matter tensor can be written in terms of the motion of the particle of matter occupying the point x^μ:

$$T^{00} = \rho_0 \frac{dx^0}{ds}\frac{dx^0}{ds} = c^2\rho_0\left(\frac{dt}{ds}\right)^2 \tag{10.8}$$

Consider now a *particle* of matter in the field which is moving with some three-dimensional velocity $\mathbf{v}$. Since we may use its time coordinate as a trajectory parameter instead of s, and since our metric at present is

$$ds^2 = c^2\,dt^2 - (dx^2 + dy^2 + dz^2) = c^2\,dt^2\left(1 - \frac{v^2}{c^2}\right) \tag{10.9}$$

we have

$$\frac{ds}{dt} = c\left(1 - \frac{v^2}{c^2}\right)^{1/2} \tag{10.10}$$

If we call $(1 - v^2/c^2)^{-1/2} = \gamma$ (which is always ≥ 1), we can write dt/ds as γ/c; the component T^{00} can then be expressed concisely as

$$T^{00} = \gamma^2\rho_0 \tag{10.11}$$

This has a simple physical interpretation; in special relativity the mass of a volume of moving material increases by a factor γ over its rest mass, while a moving three-dimensional volume element appears to have decreased in volume by the same factor. Thus, from the point of view of a fixed observer, the *density increases* by a factor γ^2. Hence, if a field of material of proper density ρ_0 flows past a fixed observer at a velocity $\mathbf{v}$, the observer will measure a density $\rho = \gamma^2\rho_0$. The component c^2T^{00} may therefore be interpreted as the *relativistic energy density* of the matter field since the only contribution to the energy of the field is due to motion of the matter.

The other components of $T^{\mu\nu}$ are also easily interpreted. Let us consider the space coordinates (x,y,z) of a particle in the field to be functions of the zeroth coordinate, time. Then the mixed space-time components are

$$T^{0i} = \rho_0 \frac{dx^0}{ds}\frac{dx^i}{ds} = \rho_0 c\left(\frac{dt}{ds}\right)\left(\frac{dx^i}{dt}\right)\left(\frac{dt}{ds}\right) \tag{10.12}$$

If we denote the three-dimensional velocity dx^i/dt by v^i, this gives

$$T^{0i} = \rho_0\gamma^2 \frac{v^i}{c} = \rho\frac{v^i}{c} \tag{10.13}$$

Similarly, the space-space components of $T^{\mu\nu}$ are given by

$$T^{ij} = \rho_0 \frac{dx^i}{ds}\frac{dx^j}{ds} = \rho_0\gamma^2\frac{v^iv^j}{c^2} = \rho\frac{v^iv^j}{c^2} \tag{10.14}$$

so the entire matter tensor can be displayed as

$$T^{\mu\nu} = \rho\begin{pmatrix} 1 & v_x/c & v_y/c & v_z/c \\ v_x/c & v_x^2/c^2 & v_xv_y/c^2 & v_xv_z/c^2 \\ v_y/c & v_yv_x/c^2 & v_y^2/c^2 & v_yv_z/c^2 \\ v_z/c & v_zv_x/c^2 & v_zv_y/c^2 & v_z^2/c^2 \end{pmatrix} \tag{10.15}$$

Let us note at this point that, just as the metric tensor corresponds to the classical gravitational *potential* according to the approximate relation $g_{00} \cong 1 + 2\varphi/c^2$, so the energy-momentum tensor corresponds to the *density* of energy in space according to the approximate relation $T^{00} \cong \rho_0$. In both cases the validity of the approximation depends upon weak fields and low velocities, $|\mathbf{v}| \ll c$.

The matter tensor can be used to write the special relativistic equations

of force-free motion for a matter field in a very elegant way. To demonstrate this let us compute the zeroth component of the divergence of $T^{\mu\nu}$:

$$T^{0\nu}{}_{|\nu} = \frac{1}{c}\frac{\partial\rho}{\partial t} + \frac{1}{c}\left[\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z}\right] \tag{10.16}$$

In three-dimensional vector notation this is

$$T^{0\nu}{}_{|\nu} = \frac{1}{c}\left[\frac{\partial\rho}{\partial t} + \nabla\cdot\rho\mathbf{v}\right] \tag{10.17}$$

The right side of this equation is familiar from the continuity equation of classical hydrodynamics:

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\rho\mathbf{v} = 0 \tag{10.18}$$

This well-known kinematic relation expresses quite generally the conservation of a quantity of material with density ρ moving with a velocity field $\mathbf{v}$. Here it expresses the conservation of matter in the sense of special relativity, which is the same as the conservation of energy. It follows that the *conservation of energy* in a free-flowing matter field is expressible as

$$T^{0\nu}{}_{|\nu} = 0 \tag{10.19}$$

The next term of the divergence of $T^{\mu\nu}$ is

$$\begin{aligned} T^{1\nu}{}_{|\nu} &= \frac{1}{c^2}\left[\frac{\partial(\rho v_x)}{\partial t} + \frac{\partial(\rho v_x^2)}{\partial x} + \frac{\partial(\rho v_xv_y)}{\partial y} + \frac{\partial(\rho v_xv_z)}{\partial z}\right] \\ &= \frac{\rho}{c^2}\left[\frac{\partial v_x}{\partial t} + v_x\frac{\partial v_x}{\partial x} + v_y\frac{\partial v_x}{\partial y} + v_z\frac{\partial v_x}{\partial z}\right] \\ &\quad + \frac{v_x}{c^2}\left[\frac{\partial\rho}{\partial t} + \frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z}\right] \end{aligned} \tag{10.20}$$

The second term of this expression is zero by the conservation-of-energy equation (10.18); the remainder can then be written as

$$T^{1\nu}{}_{|\nu} = \frac{\rho}{c^2}\left(\frac{\partial v_x}{\partial t} + \mathbf{v}\cdot\nabla v_x\right) \tag{10.21}$$

Similarly, the remaining terms of the divergence may be included in

$$T^{i\nu}{}_{|\nu} = \frac{\rho}{c^2}\left[\frac{\partial v^i}{\partial t} + \mathbf{v}\cdot\nabla v^i\right] \tag{10.22}$$

The right side of this expression is also familiar from hydrodynamics; the force-free motion of a field of material can be described by setting the Eulerian derivative or flow derivative equal to zero. The Eulerian derivative of any quantity $Q(x)$ is the change of Q as it would appear to an observer following a streaming particle; that is,

$$\frac{DQ}{Dt} \equiv \frac{\partial Q}{\partial t} + \frac{\partial Q}{\partial x^i} v^i = \frac{\partial Q}{\partial t} + \mathbf{v} \cdot \nabla Q \tag{10.23}$$

For the present case we therefore have, by setting the Eulerian derivative of the matter flow equal to zero,

$$\frac{Dv^i}{Dt} = \frac{\partial v^i}{\partial t} + \mathbf{v} \cdot \nabla v^i = 0 \tag{10.24}$$

It then follows that the equations of force-free hydrodynamic flow of the matter field can be written as

$$T^{i\nu}{}_{|\nu} = 0 \tag{10.25}$$

The Euler equations (10.24) are derived from and are equivalent to the principle of conservation of momentum. Thus, from (10.19) and (10.25), we see that demanding that the energy-momentum tensor have zero divergence is equivalent to demanding conservation of energy and conservation of momentum in the matter field.

We have described the motion of incoherent matter by the tensor law contained in (10.25) and (10.19). The mathematical advantage of this formulation comes from the fact that we can now translate this law into any coordinate system. It must take the form

$$T^{\mu\nu}{}_{\|\nu} = 0 \tag{10.26}$$

on purely formal grounds of covariance. Clearly, this law has been established only in the case that the matter moves in a Lorentzian space, i.e., a flat Riemann space. Here it expresses the conservation of energy and momentum of the matter field during the flow. We shall discuss the case of a general (nonflat) Riemann space after we have considered a few more examples of energy-momentum tensors in flat space.

10.2 Inclusion of Forces in $T^{\mu\nu}$

In the above paragraphs we have considered an incoherent matter field on which *no forces* act and whose particles do not interact. We shall now show what sort of modifications are to be made when an internal force such as pressure is present. Specifically, we shall consider a perfect fluid which is by definition characterized by a proper-density field $\rho_0(x)$, a four-vector velocity field of flow $u^\mu(x)$, and a scalar pressure field $p(x)$. We shall show that, by adding an appropriate term to the material energy-momentum tensor which we now denote as $M^{\mu\nu}$,

$$M^{\mu\nu} = \rho_0 u^\mu(x) u^\nu(x) \tag{10.27}$$

the effect of the internal force can be included in the framework of the preceding paragraphs; i.e., setting the divergence of an appropriate complete energy-momentum tensor equal to zero will give the correct equations of motion and express the conservation of energy.

We wish at first to work in the classical limit with low fluid velocities and low pressure, so we shall neglect terms of order v^2/c^2 and $p(v/c)$. Furthermore, we shall assume that the pressure is sufficiently small so that the elastic energy density of the fluid can be neglected in comparison with the energy due to material density. With these assumptions we can write the conservation of energy completely in terms of the proper matter density ρ_0:

$$\frac{\partial \rho_0}{\partial t} + \nabla \cdot (\rho_0 \mathbf{v}) = 0 \tag{10.28}$$

As is well known from fluid dynamics, the equations of motion [Eqs. (10.24)] now contain a volume force equal to the negative of the pressure gradient $\partial p/\partial x^i$:

$$\rho_0 \frac{Dv^i}{Dt} = \rho_0 \left(\frac{\partial v^i}{\partial t} + \mathbf{v} \cdot \nabla v^i \right) = f^i = -\frac{\partial p}{\partial x^i} \tag{10.29}$$

The meaning of this equation is evident. The volume element is accelerated by the pressure-force density $-\partial p/\partial x^i$, and an observer moving with the fluid experiences the acceleration Dv^i/Dt. Thus (10.29) is merely Newton's second law of dynamics. In the case of exterior volume forces, the term f^i would naturally have to be modified. Clearly, the equation is not yet relativistically invariant. It will be the task of the tensor formulation to adjust the approximate classical equations to the demands of relativistic covariance.

If we denote the matter tensor $\rho_0 u^\mu u^\nu$ as displayed in (10.15) by $M^{\mu\nu}$ to avoid confusion with the complete energy-momentum tensor $T^{\mu\nu}$, we can express the conservation of energy by

$$M^{0\nu}{}_{|\nu} = \frac{1}{c} \left(\frac{\partial \rho_0}{\partial t} + \nabla \cdot (\rho_0 \mathbf{v}) \right) = 0 \tag{10.30}$$

just as in the case of the incoherent matter field of the preceding example. The equations of motion (10.29), however, must now be expressed as

$$M^{i\nu}{}_{|\nu} = \frac{\rho_0}{c^2}\frac{Dv^i}{Dt} = \frac{\rho_0}{c^2}\left(\frac{\partial v^i}{\partial t} + \mathbf{v}\cdot\nabla v^i\right) = -\frac{1}{c^2}\frac{\partial p}{\partial x^i} \neq 0 \tag{10.31}$$

Thus $M^{\mu\nu}$ is not divergenceless because of the presence of the internal pressure force. To remedy this let us consider a 3×3 matrix S^{ij} with the property that

$$S^{ij}{}_{|j} = \frac{1}{c^2}\frac{\partial p}{\partial x^i} \tag{10.32}$$

Such a matrix is known as a three-dimensional stress tensor; its divergence represents a force. A stress tensor satisfying (10.32) is easily seen to be

$$S^{ij} = \frac{p}{c^2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{10.33}$$

Using this pressure stress tensor, we may write the equations of motion (10.31) as

$$M^{i\nu}{}_{|\nu} + S^{ij}{}_{|j} = 0 \tag{10.34}$$

Indeed, if we extend S^{ij} into a 4×4 matrix

$$S^{\mu\nu} = \frac{p}{c^2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{10.35}$$

we may combine the conservation-of-energy equation (10.30) and the equations of motion (10.31) into a single matrix equation

$$(M^{\mu\nu} + S^{\mu\nu})_{|\nu} = T^{\mu\nu}{}_{|\nu} = 0 \tag{10.36}$$

where $T^{\mu\nu}$ is explicitly

$$T^{\mu\nu} = \rho_0\begin{pmatrix} 1 & v_x/c & v_y/c & v_z/c \\ v_x/c & 0 & 0 & 0 \\ v_y/c & 0 & 0 & 0 \\ v_z/c & 0 & 0 & 0 \end{pmatrix} + \frac{p}{c^2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{10.37}$$

within our approximation, which represents the classical limit. Note, however, that $T^{\mu\nu}$ is clearly not a tensor since the relation between $\mathbf{v}$, p, and $T^{\mu\nu}$ is not covariant.

What we wish to do now is generalize (10.37) so that it becomes a true tensor. The material term $M^{\mu\nu}$ raises no problem since it is already expressible as a tensor as in (10.7), so we have only to find the proper form for $S^{\mu\nu}$. This term can be extended into an actual tensor by noting that there are only two second-rank symmetric tensor fields associated with the fluid, $g^{\mu\nu}$ and $u^\mu u^\nu$. Thus the extended $S^{\mu\nu}$ must be a linear combination of the form

$$S^{\mu\nu} = \frac{p}{c^2}[\alpha u^\mu u^\nu + \beta g^{\mu\nu}] \tag{10.38}$$

which has to reduce to (10.37) for low velocities and pressure. It is easily seen that the neglect of v^2/c^2 and $p(v/c)$ terms in (10.38) leads to

$$S^{\mu\nu} = \frac{p}{c^2}\left[\alpha\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \beta\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\right] \tag{10.39}$$

so the choice $\alpha = 1$, $\beta = -1$ evidently completes the task of generalizing $S^{\mu\nu}$ to a true tensor; we are thus led to

$$S^{\mu\nu} = \frac{p}{c^2}(u^\mu u^\nu - g^{\mu\nu}) \tag{10.40}$$

and the complete energy-momentum tensor

$$T^{\mu\nu} = \rho_0 u^\mu u^\nu + \frac{p}{c^2}(u^\mu u^\nu - g^{\mu\nu}) \tag{10.41}$$

By correspondence with the classical matrix equation (10.36) we assume that this complete energy-momentum tensor has zero divergence in a flat Riemann space:

$$T^{\mu\nu}{}_{||\nu} = 0 \tag{10.42}$$

This elegant formula is the covariant formulation of the flow of a fluid under the effect of its own internal pressure force.

In order to understand better the significance of the terms in the complete tensor $T^{\mu\nu}$, let us integrate the equation $T^{0\nu}{}_{||\nu} = 0$ over a fixed part of three-space V with boundary Σ. We have, in the usual coordinates of special relativity for which covariant and ordinary differentiation are identical,

(10.43) $$\int_V T^{0\nu}{}_{|\nu}\, dV = \frac{1}{c}\frac{\partial}{\partial t}\int_V T^{00}\, dV + \int_\Sigma T^{0i} n_i\, d\sigma = 0$$

Here n_i is the exterior normal three-vector to the surface element $d\sigma$ of Σ. This relation is in the form of a conservation law. If we recall that, in the case of incoherent matter flow, the term $c^2M^{00} = c^2\rho$ represented the energy density and $cM^{0i} = \rho v^i$ represented the momentum density, it is natural to identify the corresponding terms of the $S^{\mu\nu}$ tensor as follows:

(10.44) $$c^2S^{00} = p(\gamma^2 - 1) = p\frac{v^2}{c^2}\left(1 - \frac{v^2}{c^2}\right)^{-1}$$

is the pressure energy-density of the fluid flow, which is in general a small quantity. Similarly, cS^{0i} is the momentum density due to the fluid pressure. The total energy density is thus $c^2(M^{00} + S^{00}) = c^2T^{00}$, and the total momentum density is $c(M^{0i} + S^{0i}) = cT^{0i}$. Equation (10.43) expresses the fact that energy changes in V are caused by transport of momentum through the boundary Σ. It is convenient, therefore, to display the entire energy-momentum tensor as

(10.45) $$T^{\mu\nu} = M^{\mu\nu} + \begin{pmatrix} h & \mathbf{g} \\ \mathbf{g} & S \end{pmatrix}$$

where c^2h and $c\mathbf{g}$ represent the energy density and momentum densities associated with the pressure, and S is the three-dimensional stress tensor.

The value of the tensor law (10.42) is twofold. First, we have obtained the relativistic laws of fluid dynamics in a Lorentz covariant form which differs for high pressures and velocities from the classical noncovariant form. Second, we have a tensor law which automatically holds in all curvilinear coordinate systems. Such an elegant situation is highly desirable.

10.3 The Electromagnetic Field and $T^{\mu\nu}$

As another example involving internal forces, let us consider a flowing field of charged matter which is described by a proper density $\rho_0(x)$, a four-vector velocity u^μ, and a proper electric charge density $\sigma_0(x)$. We are still working in the flat Riemann space of special relativity, so by using the usual coordinates ct, x, y, z, we can write Maxwell's equations as

(10.46) $$F^{\mu\nu}{}_{|\nu} = s^\mu \qquad \{F_{\mu\nu|\lambda}\} = 0$$

The source vector s^μ is associated with the motion and charge of the matter field; it is in fact related to the charge density and four-velocity by

(10.47) $$s^\mu = \sigma_0 u^\mu$$

This relation is easily verified if we consider the space coordinates x^i of a particle of matter which occupies the point x^μ and moves with the flow. The x^i can be considered to be functions of the time coordinate ct, for then we may write $\sigma_0 u^\mu$ as

(10.48) $$\sigma_0\frac{dx^\mu}{ds} = \sigma_0 c\frac{dt}{ds}\left(1, \frac{1}{c}\frac{dx}{dt}, \frac{1}{c}\frac{dy}{dt}, \frac{1}{c}\frac{dz}{dt}\right) = \sigma_0 c\frac{dt}{ds}\left(1, \frac{\mathbf{v}}{c}\right)$$

Using the relation (10.10) between dt and ds for a moving particle of matter, we have

(10.49) $$\sigma_0\frac{dx^\mu}{ds} = \left(\gamma\sigma_0, \gamma\sigma_0\frac{\mathbf{v}}{c}\right)$$

It is well known from special relativity that, because of the shrinking of a moving volume element by a factor γ, the charge density σ measured by a stationary observer is increased by a factor γ; that is, $\sigma = \gamma\sigma_0$. Using this fact and noting that the convection current of the charged matter field is $\mathbf{j} = \sigma\mathbf{v}$, we can write (10.49) as

(10.50) $$\sigma_0\frac{dx^\mu}{ds} = \left(\sigma, \frac{\mathbf{j}}{c}\right)$$

which is by definition the source vector s^μ [Eq. (4.44)]. This verifies (10.47) and allows us to write the Maxwell equations (10.46) as

(10.51) $$F^{\mu\nu}{}_{|\nu} = \sigma_0 u^\mu \qquad \{F_{\mu\nu|\lambda}\} = 0$$

These equations completely determine the behavior of the electromagnetic field in terms of the behavior of the charged matter.

Let us next consider the converse problem, the behavior of the charged matter under the influence of the electromagnetic field. For convenience we shall begin working in the classical limit of small matter-velocity and neglect terms of order v^2/c^2; we shall also consider the charge density σ_0 to be sufficiently small so that we may neglect the density of electromagnetic energy. Under these assumptions the conservation-of-energy equation involves only the proper matter density ρ_0:

$$\frac{\partial \rho_0}{\partial t} + \nabla \cdot \rho_0 \mathbf{v} = 0 \tag{10.52}$$

This may also be expressed using the material tensor $M^{\mu\nu}$ as

$$M^{0\nu}{}_{|\nu} = 0 \tag{10.53}$$

precisely as in the case of the perfect-fluid example. The equations of motion of the charged material are

$$\rho_0 \frac{Dv^i}{Dt} = \rho_0 \left\{ \frac{\partial v^i}{\partial t} + \mathbf{v} \cdot \nabla v^i \right\} = f^i \tag{10.54}$$

where f^i is the Lorentz force:

$$f^i = \sigma \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{H} \right)^i \tag{10.55}$$

It can be easily verified that f^i may be expressed in terms of the tensor $F^{\mu\nu}$ as

$$f^i = -\sigma_0 F^{i\nu} u_\nu \tag{10.56}$$

where, because of the Lorentz metric, $u_0 = u^0$ and $u_i = -u^i$. To show this we write out the component $i = 1$; since $u^\nu = dx^\nu/ds$ and $u^i = u^0 v^i/c$, where v^i is of course the ordinary velocity, we have

$$-\sigma_0 F^{1\nu} u_\nu = -\sigma_0 u_0 \left(F^{10} - F^{12} \frac{v_y}{c} - F^{13} \frac{v_z}{c} \right) \tag{10.57}$$

Let us note that $u_0 = dx_0/ds = \gamma$ from (10.10), and the Lorentz volume-contraction factor is γ, so the relativistic charge density is $\sigma = \gamma\sigma_0$. Using this fact and the explicit form (4.48) for $F^{\mu\nu}$, we have

$$-\sigma_0 F^{1\nu} u_\nu = \sigma_0 \gamma \left(E_x + H_z \frac{v_y}{c} - H_y \frac{v_z}{c} \right) = \sigma \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{H} \right)_x \tag{10.58}$$

Analogous formulas hold for the other components, which verifies (10.56). Thus we may write the equations of motion of the charged matter as

$$\rho_0 \frac{Dv^i}{Dt} = -\sigma_0 F^{i\nu} u_\nu \tag{10.59}$$

or by virtue of the identity (10.22),

$$M^{i\nu}{}_{|\nu} = -\frac{\sigma_0}{c^2} F^{i\nu} u_\nu \qquad M^{\mu\nu} = \rho_0 u^\mu u^\nu \tag{10.60}$$

By analogy with the analysis of the perfect fluid, we therefore wish to obtain a matrix S with the property

$$S^{i\nu}{}_{|\nu} = \frac{\sigma_0}{c^2} F^{i\nu} u_\nu \tag{10.61}$$

so that $M^{i\nu} + S^{i\nu}$ will be divergenceless by virtue of (10.60). In order to do this we replace $i = 1, 2, 3$ by $\mu = 0, 1, 2, 3$ and lower the index in (10.61). This gives

$$S_\mu{}^\nu{}_{|\nu} = \frac{\sigma_0}{c^2} F_\mu{}^\nu u_\nu \tag{10.62}$$

Allowing $\mu = 0$ in this, we should violate the conservation law (10.52), but for small velocities and fields, this is a negligible correction. For large fields, we expect an electromagnetic correction to the energy density.

By using (10.51) we can substitute $(1/\sigma_0)F_\nu{}^\lambda{}_{|\lambda}$ for u_ν to rewrite (10.62) in the form

$$c^2 S_\mu{}^\nu{}_{|\nu} = F_\mu{}^\nu F_\nu{}^\lambda{}_{|\lambda} \tag{10.63}$$

Evidently, then, $S_\mu{}^\nu$ must be bilinear in the tensor $F^{\mu\nu}$; the most general bilinear tensor form of this kind is

$$c^2 S_\mu{}^\nu = A F_{\mu\alpha} F^{\alpha\nu} + B g_\mu{}^\nu F^{\alpha\beta} F_{\alpha\beta} \tag{10.64}$$

where A and B are constants. To determine the constants A and B, we take the divergence of $S_\mu{}^\nu$ and obtain

$$c^2 S_\mu{}^\nu{}_{|\nu} = A F_{\mu\alpha} F^{\alpha\nu}{}_{|\nu} + A F_{\mu\alpha|\nu} F^{\alpha\nu} + 2B g_\mu{}^\nu F^{\alpha\beta} F_{\alpha\beta|\nu} \tag{10.65}$$

Relabeling dummy indices and rearranging terms, we have

$$c^2 S_\mu{}^\nu{}_{|\nu} = A F_\mu{}^\nu F_\nu{}^\lambda{}_{|\lambda} + F^{\alpha\beta}(A F_{\mu\alpha|\beta} + 2B F_{\alpha\beta|\mu}) \tag{10.66}$$

This will be the desired result (10.63) if we set $A = 1$ and choose B so that the second term is zero. This is easily done, for the second term may be written

$$(10.67)\quad F^{\alpha\beta}F_{\mu\alpha|\beta} + 2BF^{\alpha\beta}F_{\alpha\beta|\mu} = \tfrac{1}{2}F^{\alpha\beta}F_{\mu\alpha|\beta} + \tfrac{1}{2}F^{\beta\alpha}F_{\mu\beta|\alpha} + 2BF^{\alpha\beta}F_{\alpha\beta|\mu}$$
$$= \tfrac{1}{2}F^{\alpha\beta}(F_{\mu\alpha|\beta} - F_{\mu\beta|\alpha} + 4BF_{\alpha\beta|\mu})$$

If we choose $B = \frac{1}{4}$, the expression in the parenthesis becomes

$$(10.68)\quad \tfrac{1}{3}(F_{\mu\alpha|\beta} - F_{\mu\beta|\alpha} + F_{\alpha\beta|\mu}) = \{F_{\mu\alpha|\beta}\}$$

which is zero, by virtue of the second set of Maxwell's equations (10.46). Therefore the choice $A = 1$, $B = \frac{1}{4}$ gives an appropriate $S_\mu{}^\nu$:

$$(10.69)\quad c^2S_\mu{}^\nu = F_{\mu\alpha}F^{\alpha\nu} + \tfrac{1}{4}g_\mu{}^\nu F^{\alpha\beta}F_{\alpha\beta}$$

The complete matrix $T^{\mu\nu} = M^{\mu\nu} + S^{\mu\nu}$ is thus

$$(10.70)\quad T^{\mu\nu} = \rho_0 u^\mu u^\nu + \frac{1}{c^2}(F^\mu{}_\alpha F^{\alpha\nu} + \tfrac{1}{4}g^{\mu\nu}F^{\alpha\beta}F_{\alpha\beta})$$

There is clearly no problem at all involved in generalizing (10.70) to a true four-dimensional tensor as there was with the previous example of the perfect fluid. As with the perfect fluid, we can interpret c^2S^{00} as the energy density of the electromagnetic field; a brief calculation using the explicit form (4.48) for $F^{\mu\nu}$ yields

$$(10.71)\quad c^2S^{00} = \frac{\mathbf{E}^2 + \mathbf{H}^2}{2}$$

which agrees with the familiar expression for electromagnetic energy density used in elementary electrodynamics. Similarly, the ith component of the momentum density of the electromagnetic field may be identified with cS^{0i}, which, by virtue of (10.69), is

$$(10.72)\quad cS^{0i} = \frac{1}{c}(\mathbf{E}\times\mathbf{H})^i$$

The vector $\mathbf{E}\times\mathbf{H}$ is the familiar Poynting vector of electrodynamics which represents momentum density, so we again have consistent agreement with classical electrodynamics. In summary, we may write the complete divergenceless energy-momentum tensor of matter field plus electromagnetic field as

$$(10.73)\quad T^{\mu\nu} = M^{\mu\nu} + \begin{pmatrix} h & \mathbf{g} \\ \mathbf{g} & S \end{pmatrix}$$

where c^2h and $c\mathbf{g}$ are the energy and momentum densities of the electromagnetic field. Note the similarity between this and the fluid tensor (10.45).

Finally, let us note that the electromagnetic field represented by $S^{\mu\nu}$ may exist in the absence of any charged material, in which case it is evident that $S^{\mu\nu}$ is divergenceless. (The reader may check that this follows directly from Maxwell's equations in vacuum.) Thus the tensor $T^{\mu\nu} = S^{\mu\nu}$ is appropriate to describe a *free electromagnetic field.*

We have again in the statement that the energy-momentum tensor is divergenceless, $T^{\mu\nu}{}_{|\nu} = 0$, an elegant formulation of the interaction between electromagnetic fields and matter. It may easily be generalized to an arbitrary coordinate system, as before, by writing in covariant tensor form

$$(10.74)\quad T^{\mu\nu}{}_{\|\nu} = 0$$

Using the preceding three examples in the framework of special theory relativity as a guide, we now make the following assumption: By including all significant physical quantities in the complete energy-momentum tensor, i.e., matter, fluid pressure, electromagnetic fields, etc., we obtain a *zero-divergence tensor* in flat space. If any quantity is omitted, it manifests itself as a force and the energy-momentum tensor cannot be considered complete and does not have a zero divergence. According to this view, physical quantities influence each other by exchanging energy and momentum in such a way as to keep the divergence of $T^{\mu\nu}$ equal to zero; i.e., total energy and momentum are conserved. We conclude that $T^{\mu\nu}$ concisely characterizes the nongravitational energy content of space. As such it is evidently a natural choice for the "tensor representing energy content of space" term in the symbolic gravitational equation (10.6). The choice of this tensor is further motivated by noting: (1) The T^{00} component of the matter tensor is ρ, the analogue of the right side of Poisson's equation (10.3). (2) The three examples we considered, (10.7), (10.41), and (10.70), involve tensors which are symmetric and, of course, second-rank, like the contracted Riemann tensor which appears in the free-space field equations. Thus we assume that the gravitational field equations have the form

$$(10.75)\quad \text{Properties of space geometry} = T^{\mu\nu}$$

where $T^{\mu\nu}$ is divergenceless and encompasses all physical quantities, except gravity, that contribute to the energy content of space. The physical interpretation of $T^{\mu\nu}{}_{\|\nu} = 0$ in a *curved* Riemann space will be taken up in the next chapter.

10.4 The Field Equations in Nonempty Space

In the previous section we motivated a choice for the term of the symbolic gravitational equation (10.6), which describes the energy content of space. In this section we must obtain a suitable tensor to describe the geometry of space in the presence of an energy field.

The simplest choice for the geometric term might seem to be the contracted Riemann tensor, which we used in the free-space field equations of Chap. 5. The field equations would then be

$$R_{\eta\gamma} = (\text{const})T_{\eta\gamma} \tag{10.76}$$

There is, however, a fatal objection to these equations, for $T_{\eta\gamma}$ has zero divergence (if all physically significant quantities are taken into account), while $R_{\eta\gamma}$ does not, as we found in (5.125). Recall, however, that in Sec. 5.8 we also expressed the free-space field equations in terms of the *divergenceless* Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R^{\alpha}{}_{\alpha} = 0 \tag{10.77}$$

It thus appears that the Einstein tensor is appropriate for use in the field equations. In fact, it can be shown that the most general second-rank tensor $B^{\alpha\gamma}$, which has zero divergence, and is constructed entirely from the metric tensor and its first and second derivatives, and is linear in the second derivatives, is a linear combination of the Einstein tensor and the metric tensor:

$$B^{\alpha\gamma} = G^{\alpha\gamma} + \Lambda g^{\alpha\gamma} \tag{10.78}$$

where Λ is an arbitrary constant. For a proof of this see Cartan (Cartan, 1922). We thus take as gravitational equations for nonempty space

$$G^{\alpha\gamma} + \Lambda g^{\alpha\gamma} = (\text{const})T^{\alpha\gamma} \tag{10.79}$$

These equations were first proposed in this form by Einstein in 1917 (Einstein, 1917).

We noted earlier that several properties are demanded of the gravitational equations. One of the requirements is that they reduce to the free-space field equations when $T^{\alpha\gamma}$, the density of energy in space, is zero. This clearly requires that the constant Λ which appears in (10.79) be zero. Actually, this requirement may be relaxed somewhat without a contradiction with experience; in Chap. 11 we shall consider the consequences of allowing Λ to be a small nonzero constant. But for now we shall assume that Λ is zero. Thus we postulate the field equations

$$G^{\alpha\gamma} = (R^{\alpha\gamma} - \tfrac{1}{2}g^{\alpha\gamma}R) = CT^{\alpha\gamma} \qquad C = \text{const} \tag{10.80}$$

By a small amount of manipulation these equations can be put into an alternative form. Contracting indices in (10.80), we have

$$R^{\alpha}{}_{\alpha} - \tfrac{1}{2}g^{\alpha}{}_{\alpha}R = CT^{\alpha}{}_{\alpha} \tag{10.81}$$

Thus

$$R = -CT^{\alpha}{}_{\alpha} = -CT \tag{10.82}$$

where T is the scalar $T^{\alpha}{}_{\alpha}$, which we shall refer to as the Laue scalar. Using this result we can write the field equations as

$$R^{\alpha\gamma} = C(T^{\alpha\gamma} - \tfrac{1}{2}g^{\alpha\gamma}T) \tag{10.83}$$

Note the symmetry between $T^{\alpha\gamma}$ and $R^{\alpha\gamma}$ evident in (10.80) and (10.83).

In the next section we shall investigate the classical limit of these equations for the case of weak fields and small velocities.

10.5 Classical Limit of the Gravitational Equations

We wish to show in this section that the field equations (10.80) are, as we desired, a generalization of Poisson's classical field equation (10.3). Besides being a validity check on the field equations, the reduction to the classical limit will give us as a by-product the value of the constant C.

Let us consider a field of matter with low proper density, moving at low velocity. The energy-momentum tensor for this situation is obtainable from the special relativistic matter tensor (10.15) if we neglect terms of order $(v/c)^2$ and $\rho_0(v/c)$:

$$T^{\mu\nu} = \begin{pmatrix} \rho_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{10.84}$$

We shall assume the flow to be stationary and therefore expect the metric to be time-independent. Using the coordinates of special relativity ct, x, y, and z, we consider a time-independent metric which is the sum of the Lorentz metric and a small time-independent perturbation $\epsilon\gamma_{\mu\nu}$,

(10.85) $$g_{\mu\nu} = \eta_{\mu\nu} + \epsilon\gamma_{\mu\nu}$$

If we neglect terms of order $\epsilon\rho_0$, the Laue scalar $T^\mu{}_\mu$ is

(10.86) $$T^\mu{}_\mu = \operatorname{Tr}\begin{pmatrix} \rho_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \rho_0$$

and the right side of the field equations is to first order in all the small quantities ρ_0, v/c, $\epsilon\gamma_{\mu\nu}$:

(10.87) $$C(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T) = C\left\{\begin{pmatrix} \rho_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} \rho_0 & 0 & 0 & 0 \\ 0 & -\rho_0 & 0 & 0 \\ 0 & 0 & -\rho_0 & 0 \\ 0 & 0 & 0 & -\rho_0 \end{pmatrix}\right\} = \frac{C\rho_0}{2}\,\delta_{\mu\nu}$$

In Sec. 9.1 we found that neglecting second-order terms in $\epsilon\gamma_{\mu\nu}$ gives the following approximate form for the contracted Riemann tensor:

(10.88) $$R_{\mu\nu} \cong \tfrac{1}{2}[\ln(-g)]_{|\mu|\nu} - \begin{Bmatrix} \beta \\ \mu\ \nu \end{Bmatrix}_{|\beta}$$

Thus the approximate field equations may be expressed as

(10.89) $$\tfrac{1}{2}[\ln(-g)]_{|\mu|\nu} - \begin{Bmatrix} \beta \\ \mu\ \nu \end{Bmatrix}_{|\beta} = \frac{C\rho_0}{2}\,\delta_{\mu\nu}$$

Consider first the case $\mu = \nu = 0$. Since we are considering a time-independent metric, the first term of (10.89) is zero, so we are left with the equation

(10.90) $$\begin{Bmatrix} \beta \\ 0\ 0 \end{Bmatrix}_{|\beta} = (g^{\alpha\beta}[00,\alpha])_{|\beta} = -C\,\frac{\rho_0}{2}$$

The Christoffel symbol of the first kind is defined by

(10.91) $$[00,\alpha] = \tfrac{1}{2}(g_{0\alpha|0} + g_{\alpha 0|0} - g_{00|\alpha})$$

Since the Lorentz metric is constant in space and time, this simplifies to

(10.92) $$[00,\alpha] = -\frac{\epsilon}{2}\gamma_{00|\alpha}$$

Furthermore, $\gamma_{\mu\nu}$ is time-independent, so [00,0] is zero. Neglecting second-order terms in $\epsilon\gamma_{\mu\nu}$, we then have

(10.93) $$g^{\beta\alpha}[00,\alpha] = \frac{\epsilon}{2}\gamma_{00|\beta}$$

which is zero for $\beta = 0$. Substituting this in (10.90), we obtain an approximate field equation for γ_{00}:

(10.94) $$\epsilon\sum_{\beta=0}^{3}\gamma_{00|\beta|\beta} = -C\rho_0$$

or by virtue of time independence,

(10.95) $$\epsilon\sum_{i=1}^{3}\gamma_{00|i|i} = -C\rho_0$$

Equation (10.95) is seen to be precisely Poisson's equation (10.3) if we make the identification

(10.96) $$-\frac{\epsilon\gamma_{00}}{C} = \frac{\varphi}{4\pi\kappa}$$

We therefore have established that the classical theory is the limiting case of the time-independent relativistic theory.

If we combine Eq. (10.96) with the result of Sec. 4.3, which relates the classical potential to the metric perturbation according to

(10.97) $$\varphi = \frac{c^2}{2}\epsilon\gamma_{00}$$

we find that

(10.98) $$C = -\frac{8\pi\kappa}{c^2}$$

Thus, by the postulate that the field equations possess the classical equations as a limit case, we have identified the constant of the field equations.

For completeness the reader may check that the entire system of Eqs.

(10.89) is satisfied by the metric tensor

$$g_{\mu\nu} = \begin{pmatrix} 1+\epsilon\gamma_{00} & & & \\ & -1+\epsilon\gamma_{00} & & \\ & & -1+\epsilon\gamma_{00} & \\ & & & -1+\epsilon\gamma_{00} \end{pmatrix} \tag{10.99}$$

giving an approximate line element of the form

$$ds^2 = \left(1 + \frac{2\varphi}{c^2}\right) c^2\, dt^2 - \left(1 - \frac{2\varphi}{c^2}\right) d\sigma^2 \tag{10.100}$$

which is of the same form as the approximate free-space line element (9.44) for a spherically symmetric field.

The most important result of this section is the identification of the constant C. We reiterate the Einstein field equations in their explicit form

$$G^{\alpha\gamma} = R^{\alpha\gamma} - \tfrac{1}{2}g^{\alpha\gamma}R = -\frac{8\pi\kappa}{c^2}\,T^{\alpha\gamma} \tag{10.101a}$$

or equivalently,

$$R^{\alpha\gamma} = -\frac{8\pi\kappa}{c^2}\,(T^{\alpha\gamma} - \tfrac{1}{2}g^{\alpha\gamma}T) \tag{10.101b}$$

Exercises

10.1 Show that (10.99) is a solution of the approximate field equations (10.89).

10.2 Discuss the general form that the energy-momentum tensor for a scalar field may take, following the discussion of $T_{\mu\nu}$ for the electromagnetic field.

10.3 Repeat the above exercise for a vector field.

10.4 As in Exercise 4.2, define a complex tensor for the electromagnetic field by

$$\omega_{\mu\nu} = F_{\mu\nu} + i({}^*F_{\mu\nu})$$

and denote its complex conjugate by $\bar{\omega}_{\mu\nu}$. Show that $\omega_{\mu\nu}$ is related to

the energy-momentum tensor of the electromagnetic field by

$$-\frac{1}{2c^2}\,\mathrm{Re}\,(\bar{\omega}^{\alpha}{}_{\gamma}\omega_{\alpha\beta}) = S_{\gamma\beta}$$

10.5 Recall from Exercise 4.3 that $\omega_{\alpha\beta}\bar{\omega}^{\alpha\beta} = 0$ and show from this that $S_{\gamma\beta}$ is traceless.

10.6 Perform a "duality rotation" on $F_{\mu\nu}$, defined as

$$F'_{\mu\nu} = F_{\mu\nu}\cos\theta + i({}^*F_{\mu\nu})\sin\theta$$

Show that $S_{\mu\nu}$ is unchanged by this transformation. What is the explicit effect on a simple plane wave? If $F_{\mu\nu}$ obeys Maxwell's equations, does $F'_{\mu\nu}$? What happens to the invariants discussed in Exercise 4.3?

10.7 The Petrov classification discussed in Chap. 5 relies on the vacuum field equation $R_{\mu\nu} = 0$. In order to extend the classification scheme to nonempty space we introduce the Weyl tensor, $C_{\alpha\beta\gamma\delta} = R_{\alpha\beta\gamma\delta} + g_{\alpha\gamma}R_{\delta\beta} - g_{\alpha\delta}R_{\gamma\beta} + R_{\alpha\gamma}g_{\delta\beta} - R_{\alpha\delta}g_{\gamma\beta} - \frac{1}{3}(g_{\alpha\gamma}g_{\delta\beta} - g_{\alpha\delta}g_{\gamma\beta})R$ (see Prob. 5.2 for some properties of this tensor). Show that it has the same algebraic symmetry properties as the Riemann tensor and automatically satisfies $C^{\alpha}{}_{\mu\alpha\nu} = 0$, the analogue of the vacuum field equations.

10.8 Show from the above that the Petrov classification of vacuum space-times carries over directly to nonempty space-times if the Riemann tensor is replaced by the Weyl tensor.

10.9 Show that Λ introduced in (10.79) must indeed be a constant and cannot be a function of the coordinates.

Problems

10.1 In the canonical theory of fields a Lagrangian density L is obtained which leads to the desired field equations via the Euler-Lagrange method. An explicit form for a canonical energy-momentum tensor can be obtained from L. Use the canonical formalism to obtain $T_{\mu\nu}$ for a scalar field, a general vector field, and the electromagnetic field. What of the gravitational field? Is the canonical $T_{\mu\nu}$ necessarily a symmetric tensor? What physical demand indicates that it should be symmetric? (See Bjorken and Drell, 1965.)

10.2 Consider a spherical body of constant density and total geometric mass m rotating slowly with angular frequency ω. The approximate metric can be obtained from the field equations by working to first order in m and ω. The appropriate general metric form can be inferred by means analogous to those mentioned in the text, Sec. 6.1. The result is

$$ds^2 = A^2(r)c^2\,dt^2 - B^2(r)[dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\varphi^2)] + 2\Omega(r)r^2\sin^2\theta\,d\varphi\,dt$$

Solve the field equations for the functions A, B, and Ω, using appropriate boundary conditions at the surface of the body. Thereby obtain the Lense-Thirring result used in Sec. 7.7 (see Lense and Thirring, 1918; Adams et al., 1974).

Bibliography

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

Adams, R., J. M. Cohen, R. J. Adler, and C. Sheffield (1974): Analytic Pulsar Models, *Astrophys. J.*,

Bauer, H. (1918): Kugelsymmetrische Lösungssysteme der Einsteinschen Feldgleichungen der Gravitation für eine ruhende, gravitierende Flüssigkeit mit linearer Zustandsgleichung, *Sitzber. Akad. Wiss. Wien*, Abt. IIa, **127**:2141–2227.

Bjorken, J. D., and S. D. Drell (1965): "Relativistic Quantum Fields," New York.

Cartan, E. (1922): Sur les Équations de la gravitation d'Einstein, *J. Math. Pures Appl.*, **1**:141–203.

Cohen, J. M. (1968): Angular Momentum and the Kerr Metric, *J. Math. Phys.*, **9**:905.

Einstein, A. (1915): Die Feldgleichungen der Gravitation, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 844–847.

Einstein, A. (1917): Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 142–152.

Lense, J., and H. Thirring (1918): Über den Einfluss der Eigenrotation der Zentralkörper auf de Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie, *Phys. Z.*, **19**:156.

Weyl, H. (1919): Über die statischen kugelsymmetrischen Lösungen von Einsteins "kosmologischen" Gravitationsgleichungen, *Phys. Z.*, **20**:31–34.

CHAPTER

11

Further Consequences of the Field Equations

This chapter will be a continuation of the investigation of the field equations which we began in Chap. 10. We shall deal with more formal and less completely solved questions than the simple physical problems of the last chapter. The two topics to be discussed are the relation of the equations of motion to the field equations and the nature of general relativistic conservation laws. Neither subject can be considered complete or closed at present, so this chapter serves only as a short introduction to the current work on these problems. Finally we shall briefly discuss the modification of relativity theory proposed by Brans and Dicke.

11.1 The Equations of Motion

Up to now the procedure in determining the motion of a test particle in a given physical situation was the following: (1) We described the distribution of matter and fields by means of the energy-momentum tensor. (2) We calculated the metric field from the Einstein field equations by integrating the Ricci tensor. (3) We found the trajectory of the test particle as a geodesic of the Riemannian geometry. In this method, we thus used two different basic laws: (*a*) The Einstein field equations and (*b*) the postulate of geodesic motion. However, these two basic laws cannot be independent of each other. The test particle considered is a part of the total matter which enters into the energy-momentum tensor and has been split off unnaturally in order to be studied with greater convenience. The Einstein field equations lead to certain differential equations for the energy-momentum tensor which determine its behavior in time and space. Hence, in particular, one should expect that the motion of a test particle should be somehow contained in the field equations. In other words, it seems possible that the postulate of geodesic motion could be deduced from the field equations instead of being axiomatically required.

The first attempt to study the motion of particles from the field equations without the postulate of geodesic motion was made by Einstein and Grommer in 1927. They showed that a singularity of the metric field could not be freely prescribed in space-time but had a specific form as a consequence of the field equations (Einstein and Grommer, 1927). Clearly, a singularity of the field might be interpreted as a material point, and the interrelation of the two fundamental laws of general relativity was thus indicated. The reasoning of Einstein and Grommer was further developed by Einstein, Infeld, and their collaborators and the Russian school of Fock during the thirties and found a final exposition in the work of Infeld and Plebanski (1960).

Let us point out that the basic reason for the interrelation of the two basic laws is the nonlinear character of the field equations of general relativity. Thus we cannot simply add the effects of separate bodies and their fields to obtain a resultant field as we do in the classical theory of gravitation. Instead, we must consider the combined field as an inseparable whole. This is due to the fact that the gravitational field contains energy and must therefore serve as part of its own source, as we noted in Chap. 9. In a linear theory we might create additional solutions by adding or integrating solutions with point singularities. The nonlinear theory precludes such construction. The nonexistence of solutions with arbitrary singularities must strongly affect the dynamics of a material point since such a point may be viewed as a singularity of the field.

Indeed, we shall show in this section that the field equations actually specify unique equations of motion for the case of a point particle in a gravitational field and that the ensuing trajectory of that particle is a geodesic of the corresponding metric in agreement with our previous postulate of geodesic motion. Thus this postulate appears as a consequence of the field equations, and not as an independent axiom of the theory.

Our derivation will follow the method of Levi-Civita (1929). The reader who wishes a more detailed treatment is referred to Infeld and Plebanski (1960). We shall consider the streamlines of the particles in a cloud of dust represented by the matter tensor discussed in Sec. 10.1.

From the field equations (10.80) we are assured that the energy-momentum tensor $T_{\mu\nu}$ has zero divergence:

$$T^{\mu\nu}{}_{\|\nu} = 0 \tag{11.1}$$

Let us investigate the implications of this relation for the matter tensor (10.7), which represents a cloud of dust that does not interact with itself except via gravitation. Setting the covariant derivative of the matter tensor equal to zero, we obtain directly

$$(\rho_0 u^\nu)_{\|\nu} u^\mu + \rho_0 (u^\nu u^\mu{}_{\|\nu}) = (\rho_0 u^\nu)_{\|\nu} u^\mu + \rho_0 A^\mu = 0 \tag{11.2}$$

where A^μ is introduced for later convenience. We also know that the four-vector u^μ has unit magnitude

$$u^\alpha u_\alpha = \frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} g_{\alpha\beta} = 1 \tag{11.3}$$

With this relation we can obtain a very beautiful result from (11.2). Differentiation of (11.3) with respect to s gives

$$(u^\alpha u_\alpha)_{|\nu} u^\nu = (u^\alpha u_\alpha)_{\|\nu} u^\nu = 2(u^\alpha{}_{\|\nu} u^\nu) u_\alpha = 0 \tag{11.4}$$

We recognize this as the statement that u_α and A^α are orthogonal:

$$A^\alpha u_\alpha = 0 \tag{11.5}$$

If (11.2) is now multiplied by u_μ, we see with the aid of (11.5) that

$$(\rho_0 u^\nu)_{\|\nu} u^\mu u_\mu + \rho_0 A^\mu u_\mu = (\rho_0 u^\nu)_{\|\nu} = 0 \tag{11.6}$$

That is, the quantity $\rho_0 u^\nu$, which we may interpret as the momentum density, is conserved. This is an interesting result in itself, but it implies even more. From (11.2) we now see that the vector A^μ is identically zero:

$$A^\mu = u^\nu u^\mu{}_{\|\nu} = u^\nu \left(u^\mu{}_{|\nu} + \begin{Bmatrix} \mu \\ \alpha\ \nu \end{Bmatrix} u^\alpha \right) = 0 \tag{11.7}$$

The vector u^ν represents a streamline of the dust cloud, and if an individual dust particle is assigned coordinates x^ν, we can identify u^ν with dx^ν/ds. Then (11.7) becomes a constraint on the particle's motion.

$$\frac{dx^\nu}{ds}\frac{\partial}{\partial x^\nu}\left(\frac{dx^\mu}{ds}\right) + \begin{Bmatrix} \mu \\ \alpha\ \nu \end{Bmatrix} \frac{dx^\nu}{ds}\frac{dx^\alpha}{ds} = \frac{d^2x^\mu}{ds^2} + \begin{Bmatrix} \mu \\ \alpha\ \nu \end{Bmatrix} \frac{dx^\nu}{ds}\frac{dx^\alpha}{ds} = 0 \tag{11.8}$$

In fact the constraint is a complete equation of motion for the dust particle—the geodesic equation which we anticipated.

We have thus shown that the field equations imply a unique equation of motion for the elements of a cloud of dust particles moving under the influence of whatever gravitational field is present.

It is evident from the above that a vanishingly small globule of dust described by a field u^α that is nearly constant over the size of the globule will move on the geodesic. Such a globule serves as a good test body for the following two reasons: (1) its energy-momentum tensor contains only one scalar parameter, the density ρ_0, and is therefore as simple as possible; (2) its internal structure is described entirely by a uniform velocity field and in this sense also is as simple as possible. It should not be surprising that bodies with more complicated internal structure do not necessarily move on geodesics, e.g., extended bodies and bodies with internal motion (see Probs. 11.1 and 11.2).

11.2 Conservation Laws in General Relativity: Energy-Momentum of the Gravitational Field

We have derived the Schwarzschild line elements as one exact solution of the field equations of general relativity theory and discussed some physical implications of these results. There are very few more known exact solutions of the field equations with physical significance. In facing more complicated problems in general relativity, one is forced into approximation methods and numerical procedures. This is certainly not a new feature of relativistic mechanics, for even in classical mechanics not too many significant problems can be solved in closed analytical form. In celestial mechanics the Kepler problem plays the role analogous to the Schwarzschild problem in relativistic astronomy. It can be solved exactly, and is, at the same time, the starting point for many perturbation and approximation procedures which are needed to solve more complicated and practical problems of celestial mechanics. However, in classical mechanics we have, besides the few explicitly solved standard problems, a large number of general principles which often allow qualitative conclusions of great importance without necessitating explicit solutions of the differential equations involved. It is our aim to give somewhat similar general results in the case of the general relativity theory.

We wish to discuss first the question of conservation laws in connection with the field equations. Since the energy-momentum tensor is proportional to the Einstein tensor,

(11.9) $$G_{\mu\nu} = CT_{\mu\nu} \qquad C = -\frac{8\pi\kappa}{c^2}$$

and since the Einstein tensor has a vanishing covariant divergence, we have the important general identity

(11.10) $$T^{\mu\nu}{}_{\|\nu} = 0$$

Let us briefly review the fact that in a flat space the vanishing of the divergence of a tensor leads always to an interpretation in terms of conservation laws. In a flat space we can by definition use a coordinate system in which the Christoffel symbols vanish everywhere. Hence covariant differentiation reduces in such coordinates to ordinary differentiation. Suppose now that the tensor $T^{\mu\nu}$ is different from zero only in a finite region of space at every moment. Then the support of $T^{\mu\nu}$, that is, the part of the space-time manifold where the tensor is different from zero, can be enclosed in a "four-tube" D^4 on the wall of which $T^{\mu\nu} \equiv 0$. Consider the part Δ^4 of D^4, which is cut off by the hypersurfaces $t = t_i$ and $t = t_f$ (see Fig. 11.1). The boundary of Δ^4 consists

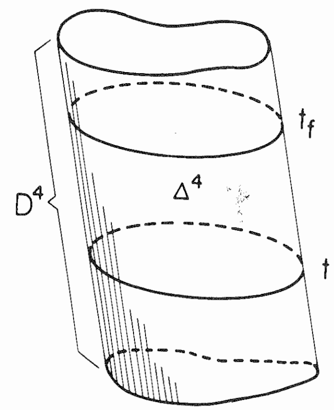


Fig. 11.1

of two three-dimensional spaces at the initial time t_i and the final time t_f and of a part of the wall of D^4 on which the tensor vanishes. Hence, by Gauss's theorem,

(11.11) $$\int_{\Delta^4} T^{\mu\nu}{}_{|\nu}\, d^4x = \int_{t_f} T^{\mu 0}\, d^3x - \int_{t_i} T^{\mu 0}\, d^3x = 0$$

Thus the vector quantity

(11.12) $$P^\mu = \int T^{\mu 0}\, d^3x$$

does not change with time if we integrate over a three-dimensional spacelike region which may change with time, but is chosen in such a way that the tensor $T^{\mu\nu}$ vanishes on its boundary. We may consider P^μ as a conserved quantity and give it a physical interpretation. If $T^{\mu\nu}$ is the energy-momentum tensor, the quantity P^μ can be interpreted as the energy-momentum vector of special relativity and $T^{\mu 0}$ can be interpreted as the density in space of this vector.

Such a straightforward consideration is not possible in a general Riemann space with nonzero Christoffel symbols. Instead, we are forced to search for an alternative conserved quantity to replace P^μ. The reason for this is clear, for in a curved space, energy and field interact, so we can expect only some combination of $T^{\mu\nu}$ and the gravitational field energy to be conserved. We shall find that an expression involving $T^{\mu\nu}$ and a function of the gravitational field variables is indeed conserved. Unfortunately, however, we shall see that there remain problems of interpretation, covariance, and uniqueness connected with the conserved quantity.

To investigate (11.10) in a general Riemann space let us first bring it into explicit form. We have, by definition of covariant differentiation,

$$T_\mu{}^\nu{}_{\|\nu} = T_\mu{}^\nu{}_{|\nu} + \begin{Bmatrix} \beta \\ \nu\ \beta \end{Bmatrix} T_\mu{}^\nu - \begin{Bmatrix} \alpha \\ \mu\ \beta \end{Bmatrix} T_\alpha{}^\beta = 0 \tag{11.13}$$

and from Eq. (3.11), we have

$$\begin{Bmatrix} \beta \\ \nu\ \beta \end{Bmatrix} = (\log \sqrt{-g})_{|\nu} = \frac{(\sqrt{-g})_{|\nu}}{\sqrt{-g}} \tag{11.14}$$

Furthermore, the symmetry of $T^{\mu\nu}$ implies that

$$\begin{Bmatrix} \alpha \\ \mu\ \beta \end{Bmatrix} T_\alpha{}^\beta = g^{\alpha\tau}[\mu\beta,\tau]T_\alpha{}^\beta = [\mu\beta,\tau]T^{\tau\beta} = \tfrac{1}{2}g_{\tau\beta|\mu}T^{\tau\beta} \tag{11.15}$$

so we can put (11.13) into simpler form,

$$T_\mu{}^\nu{}_{|\nu} + \frac{(\sqrt{-g})_{|\nu}}{\sqrt{-g}} T_\mu{}^\nu - \tfrac{1}{2}g_{\tau\beta|\mu}T^{\tau\beta} = 0 \tag{11.16}$$

that is,

$$(\sqrt{-g}\, T_\mu{}^\nu)_{|\nu} - \tfrac{1}{2}g_{\tau\beta|\mu} \sqrt{-g}\, T^{\tau\beta} = 0 \tag{11.17}$$

Because of the field equations (11.10), the last term of (11.17) may be expressed in terms of the Einstein tensor

$$\sqrt{-g}\, T_\mu{}^\nu{}_{\|\nu} = (\sqrt{-g}\, T_\mu{}^\nu)_{|\nu} - \tfrac{1}{2}g_{\tau\beta|\mu} \frac{\sqrt{-g}}{C} G^{\tau\beta} = 0 \tag{11.18}$$

(Note that this particularly simple result holds only for the mixed tensor $T_\mu{}^\nu$.) The problem is now clear; if we could write the last term of (11.18) as the ordinary divergence of some quantity $\sqrt{-g}\, t_\mu{}^\nu$, then we should have an identity which would lead to a conservation law via a simple application of Gauss's theorem, as we demonstrated at the beginning of this section.

It should be observed that we have reduced the question of conservation laws in physics to a problem of differential geometry. Indeed, the quantity $\sqrt{-g}\, t_\mu{}^\nu$ which we seek depends only on the metric considered. Once we have determined for each given metric such a quantity, we shall be able to assert conservation laws for a host of physical problems. Indeed, the different physical situations are characterized by the form of the energy-momentum tensor $T^{\mu\nu}$, while the construction which we shall now make will depend only on the structure of the geometric tensor $g^{\mu\nu}$.

The Ricci tensor is built in a very specific nonlinear way from the metric tensor field $g_{\mu\nu}$. In order to exhibit this dependence clearly, it will be useful to vary the field $g_{\mu\nu}$ arbitrarily and to study the effect of this change on the Einstein tensor. We are thus led to methods which are typically used in the calculus of variations. The following considerations will not only lead us to the $\sqrt{-g}\, t_\mu{}^\nu$ terms, but are of independent mathematical interest. They also lay the groundwork for the variational principles in general relativity theory which correspond to the well-known variational formulations in classical dynamics.

We begin our discussion with the scalar density $\mathfrak{R}$ based on the Riemann scalar R:

$$\begin{aligned} \mathfrak{R} &= \sqrt{-g}\, R = \sqrt{-g}\, g^{\sigma\rho} R_{\sigma\rho} \\ &= \sqrt{-g}\, g^{\sigma\rho} \left[\begin{Bmatrix} \alpha \\ \sigma\ \alpha \end{Bmatrix}_{|\rho} - \begin{Bmatrix} \alpha \\ \sigma\ \rho \end{Bmatrix}_{|\alpha} + \begin{Bmatrix} \beta \\ \sigma\ \alpha \end{Bmatrix} \begin{Bmatrix} \alpha \\ \rho\ \beta \end{Bmatrix} \right. \\ &\qquad \left. - \begin{Bmatrix} \alpha \\ \sigma\ \rho \end{Bmatrix} \begin{Bmatrix} \beta \\ \alpha\ \beta \end{Bmatrix} \right] \end{aligned} \tag{11.19}$$

If D^4 is an arbitrary region in four-space, we can form the invariant integral

$$J = \int_{D^4} \mathfrak{R}\, d^4x \tag{11.20}$$

We can now show that the tensor density $G_{\mu\nu}\sqrt{-g}$ can be obtained as the variational derivative of this expression under a variation of the metric tensor in D^4, which vanishes on the boundary of the region. More precisely, we consider a change of the metric tensor $\delta g_{\mu\nu}$ such that both $\delta g_{\mu\nu}$ and $\delta g_{\mu\nu|\lambda}$ vanish on the boundary of D^4.

We observe that, if we change from the marker system x^α to a marker system $\bar{x}^\alpha$, the metric tensor transforms according to

$$\bar{g}_{\mu\nu} = \frac{\partial x^\alpha}{\partial \bar{x}^\mu}\frac{\partial x^\beta}{\partial \bar{x}^\nu} g_{\alpha\beta} \tag{11.21}$$

The same law is also valid for the varied metric tensor $g_{\mu\nu} + \delta g_{\mu\nu}$. Hence the variation $\delta g_{\mu\nu}$ also transforms like a tensor if we change the coordinate system. The same is true for all tensors built from the metric tensor. In particular, the tensor $R_{\mu\nu}$ will go over into $R_{\mu\nu} + \delta R_{\mu\nu}$ under the variation of the metric tensor, and by the same argument it follows that $\delta R_{\mu\nu}$ will transform as a tensor. Moreover, the variation of the metric tensor field $\delta g_{\mu\nu}$ gives rise to an interesting variation in the Christoffel symbols $\delta \begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix}$. From the transformation law for $\begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix}$ as given in (2.5), we recognize that the variation $\delta \begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix}$ must indeed transform as a tensor,

$$\overline{\delta \begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix}} = \frac{\partial \bar{x}^\alpha}{\partial x^\kappa}\frac{\partial x^\lambda}{\partial \bar{x}^\beta}\frac{\partial x^\sigma}{\partial \bar{x}^\gamma}\, \delta \begin{Bmatrix} \kappa \\ \lambda\ \sigma \end{Bmatrix} \tag{11.22}$$

since the inhomogeneous term in (2.5) depends on the change of the coordinate systems only, but not on the metric used. We already mentioned, in Chap. 2, this important fact that the variations of connections under a change of metric transform like tensors. This will considerably simplify the variational calculations.

Following an idea of Palatini (1919), we now compute the variation of $R_{\mu\nu}$ at a given point by introducing a locally geodesic coordinate system. In such a system all Christoffel symbols vanish at the point considered, by definition. But their variations do not vanish since, with the varied metric, the coordinate system will in general no longer be locally geodesic. Since ordinary and covariant differentiation are the same in a geodesic coordinate system, we have, by definition of $R_{\mu\nu}$,

$$\delta R_{\mu\nu} = \delta \begin{Bmatrix} \alpha \\ \mu\ \alpha \end{Bmatrix}_{|\nu} - \delta \begin{Bmatrix} \alpha \\ \mu\ \nu \end{Bmatrix}_{|\alpha} = \left(\delta \begin{Bmatrix} \alpha \\ \mu\ \alpha \end{Bmatrix}\right)_{\|\nu} - \left(\delta \begin{Bmatrix} \alpha \\ \mu\ \nu \end{Bmatrix}\right)_{\|\alpha} \tag{11.23}$$

The terms $\delta R_{\mu\nu}$ and $\left(\delta \begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix}\right)_{\|\delta}$ are tensors. Hence the extreme terms of (11.23) form a tensor equation. It has been established in a convenient coordinate system, but it must hold in all coordinate systems. Thus

$$\delta R_{\mu\nu} = \left(\delta \begin{Bmatrix} \alpha \\ \mu\ \alpha \end{Bmatrix}\right)_{\|\nu} - \left(\delta \begin{Bmatrix} \alpha \\ \mu\ \nu \end{Bmatrix}\right)_{\|\alpha} \tag{11.23'}$$

is a generally valid expression for the variation of $R_{\mu\nu}$.

Next let us obtain the variation of $\sqrt{-g}$, which also occurs in the definition of J. The determinant g may be expanded in its elements of the νth column and their cofactors as

$$g = \sum_\mu{}' g_{\mu\nu}\Delta^{\mu\nu} \tag{11.24}$$

where $\Delta^{\mu\nu}$ is the cofactor of $g_{\mu\nu}$ and ν is any fixed column index. Clearly, then,

$$\frac{\partial g}{\partial g_{\mu\nu}} = \Delta^{\mu\nu} \tag{11.25}$$

and hence

$$\delta g = \frac{\partial g}{\partial g_{\mu\nu}}\,\delta g_{\mu\nu} = \Delta^{\mu\nu}\,\delta g_{\mu\nu} \tag{11.25'}$$

On the other hand, we can use the definition of the inverse matrix of $g_{\mu\nu}$ to find

$$g^{\mu\nu} = \frac{1}{g}\Delta^{\mu\nu} \tag{11.26}$$

and hence (10.40′) becomes

$$\delta g = g g^{\mu\nu}\,\delta g_{\mu\nu} \tag{11.27}$$

This result can also be formulated in terms of $\delta g^{\mu\nu}$ by noting that $g^{\mu\nu}g_{\mu\nu}$ is the trace of the invariant Kronecker tensor $g^\mu{}_\nu$ and has the value $\delta^\nu{}_\nu = 4$. Hence

$$\delta(g^{\mu\nu}g_{\mu\nu}) = g_{\mu\nu}\,\delta g^{\mu\nu} + g^{\mu\nu}\,\delta g_{\mu\nu} = 0 \tag{11.28}$$

We have remarked that $\delta g_{\mu\nu}$ is a tensor, and hence $g^{\mu\nu}\,\delta g_{\mu\nu}$ is a scalar. But observe that the variational operator δ and the operation of raising and lowering indices do not commute. Thus the contravariant form of $\delta g_{\mu\nu}$ is not $\delta g^{\mu\nu}$, as is evident from (11.28). We see, however, that

$$g^{\mu\nu}\,\delta g_{\mu\nu} = -g_{\mu\nu}\,\delta g^{\mu\nu}$$

and (11.27) becomes

$$\delta g = -g g_{\mu\nu}\,\delta g^{\mu\nu} \tag{11.29}$$

Thus, finally,

$$\delta\sqrt{-g} = -\frac{1}{2}\frac{1}{\sqrt{-g}}\,\delta g = -\tfrac{1}{2}\sqrt{-g}\,g_{\mu\nu}\,\delta g^{\mu\nu} \tag{11.30}$$

This completes the computation of the variation of the various factors in J.

The total variation of J can now be written as

$$\begin{aligned}\delta J &= \int_{D^4} \delta(\sqrt{-g}\,g^{\mu\nu}R_{\mu\nu})\,d^4x \\ &= \int_{D^4} [\sqrt{-g}\,g^{\mu\nu}\,\delta R_{\mu\nu} + g^{\mu\nu}R_{\mu\nu}\,\delta\sqrt{-g} + R_{\mu\nu}\sqrt{-g}\,\delta g^{\mu\nu}]\,d^4x\end{aligned} \tag{11.31}$$

Using the explicit forms for $\delta R_{\mu\nu}$ and $\delta\sqrt{-g}$ from (11.23′) and (11.30), we obtain

$$\begin{aligned}\delta J &= \int_{D^4}\sqrt{-g}\,g^{\mu\nu}\left[\left(\delta\begin{Bmatrix}\alpha\\ \mu\ \alpha\end{Bmatrix}\right)_{\|\nu} - \left(\delta\begin{Bmatrix}\alpha\\ \mu\ \nu\end{Bmatrix}\right)_{\|\alpha}\right]d^4x \\ &\qquad + \int_{D^4}(R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R)\sqrt{-g}\,\delta g^{\mu\nu}\,d^4x \\ &= \int_{D^4}\sqrt{-g}\,g^{\mu\nu}\left[\left(\delta\begin{Bmatrix}\alpha\\ \mu\ \alpha\end{Bmatrix}\right)_{\|\nu} - \left(\delta\begin{Bmatrix}\alpha\\ \mu\ \nu\end{Bmatrix}\right)_{\|\alpha}\right]d^4x \\ &\qquad + \int_{D^4}G_{\mu\nu}\sqrt{-g}\,\delta g^{\mu\nu}\,d^4x\end{aligned} \tag{11.32}$$

Since the covariant derivative of the metric tensor vanishes, we may rewrite the integrand of the first integral in (11.32) in the form

$$\begin{aligned}&\left[\left(g^{\mu\nu}\,\delta\begin{Bmatrix}\alpha\\ \mu\ \alpha\end{Bmatrix}\right)_{\|\nu} - \left(g^{\mu\nu}\,\delta\begin{Bmatrix}\alpha\\ \mu\ \nu\end{Bmatrix}\right)_{\|\alpha}\right]\sqrt{-g} \\ &\qquad = (v^\nu{}_{\|\nu} - w^\alpha{}_{\|\alpha})\sqrt{-g}\end{aligned} \tag{11.33}$$

where v^ν and w^α are contravariant vectors. We use now the general formula (3.12), which allows us to express the covariant divergence of a contravariant vector in the form

$$\sqrt{-g}\,v^\nu{}_{\|\nu} = (\sqrt{-g}\,v^\nu)_{|\nu} \tag{11.34}$$

in which the right side is an ordinary divergence term. Thus the entire integrand of the first integral in (11.32) is seen to be an ordinary divergence, and by partial integration we can express this integral as a surface integral over the boundary of D^4. Since $\delta g^{\mu\nu}$ and $\delta\begin{Bmatrix}\alpha\\ \beta\ \gamma\end{Bmatrix}$ vanish on this boundary, the first integral must therefore be zero. Thus the expression (11.32) for the variation of J reduces to the simple fundamental formula

$$\delta J = \int_{D^4} G_{\mu\nu}\sqrt{-g}\,\delta g^{\mu\nu}\,d^4x \tag{11.35}$$

We have thus proved that *the Einstein tensor density $G_{\mu\nu}\sqrt{-g}$ is the variational derivative of the invariant J integral over the Riemann scalar density $R\sqrt{-g}$.* We state again the fact that (11.35) is valid only for variations of the metric tensor and its first derivatives, which vanish on the boundary of D^4. For more general variations we should have to add certain surface integrals on the right-hand side.

The value of the variational formula (11.35) for formal transformations and the derivation of identities for the Einstein tensor is obvious. We have found the variation of the integral J by simple considerations of covariance. On the other hand, the integrand $\mathcal{R}$ of J is a complicated nonlinear function of the $g_{\mu\nu}$ and their first and second derivatives. Using the standard formalism of the calculus of variations (the usual Euler-Lagrange approach), we can express the variational derivative $G_{\mu\nu}$ in terms of partial derivatives of $\mathcal{R}$ with respect to its variables. Observe, however, that even the second derivatives of the varied functions $g_{\mu\nu}$ enter in $\mathcal{R}$; this will make the Euler-Lagrange terms rather complicated. It is now of great convenience that we can split $\mathcal{R}$ into one term $\mathfrak{A}$, which depends only on the $g_{\mu\nu}$ and their first derivatives, and into a term $\mathfrak{B}$, which is a linear combination of derivatives. We shall see that the contribution of the second term of the integrand $\mathcal{R}$ of J will be reducible to an integral over the boundary of the integration domain D^4. Thus the variations of J and of the integral over the much simpler term $\mathfrak{A}$ will be identical. We proceed now to the determination of $\mathfrak{A}$ and carry out the above program. We first rewrite the definition (11.19) of the Riemann density by use of the identity

$$\begin{aligned}&\sqrt{-g}\,g^{\sigma\rho}\left[\begin{Bmatrix}\alpha\\ \sigma\ \alpha\end{Bmatrix}_{|\rho} - \begin{Bmatrix}\alpha\\ \sigma\ \rho\end{Bmatrix}_{|\alpha}\right] \\ &\qquad = \left(\sqrt{-g}\,g^{\sigma\rho}\begin{Bmatrix}\alpha\\ \sigma\ \alpha\end{Bmatrix}\right)_{|\rho} - \left(\sqrt{-g}\,g^{\sigma\rho}\begin{Bmatrix}\alpha\\ \sigma\ \rho\end{Bmatrix}\right)_{|\alpha} \\ &\qquad\qquad + \begin{Bmatrix}\alpha\\ \sigma\ \rho\end{Bmatrix}(\sqrt{-g}\,g^{\sigma\rho})_{|\alpha} - \begin{Bmatrix}\alpha\\ \sigma\ \alpha\end{Bmatrix}(\sqrt{-g}\,g^{\sigma\rho})_{|\rho}\end{aligned} \tag{11.36}$$

If we substitute this relation into (11.19), we have achieved the decomposition of $R\sqrt{-g}$ into a divergence term and a term containing at most first derivatives of the $g_{\mu\nu}$. However, we can recombine terms in a very remarkable way and find an elegant expression for the term $\mathfrak{A}$.

We start with the fact that the covariant derivative of $g^{\sigma\rho}$ vanishes; this implies that

$$g^{\sigma\rho}{}_{\|\alpha} = g^{\sigma\rho}{}_{|\alpha} + \left\{{\sigma \atop \alpha\ \beta}\right\} g^{\beta\rho} + \left\{{\rho \atop \alpha\ \beta}\right\} g^{\sigma\beta} = 0 \tag{11.37}$$

This identity allows us to replace all terms $g^{\sigma\rho}{}_{|\alpha}$ by elements $g^{\mu\nu}$ and Christoffel symbols. Similarly, we may use Eq. (11.14) to express the derivatives of $\sqrt{-g}$ in terms of Christoffel symbols. Thus the last two terms in (11.36) become

$$\begin{aligned}(11.38)\quad & \left\{{\alpha \atop \sigma\ \rho}\right\} (\sqrt{-g}\,g^{\sigma\rho})_{|\alpha} - \left\{{\alpha \atop \sigma\ \alpha}\right\} (\sqrt{-g}\,g^{\sigma\rho})_{|\rho} \\ &= \left\{{\alpha \atop \sigma\ \rho}\right\}\left[-\left\{{\sigma \atop \alpha\ \beta}\right\} g^{\beta\rho} - \left\{{\rho \atop \alpha\ \beta}\right\} g^{\sigma\beta} + \left\{{\beta \atop \alpha\ \beta}\right\} g^{\sigma\rho}\right]\sqrt{-g} \\ &\quad - \left\{{\alpha \atop \sigma\ \alpha}\right\}\left[-\left\{{\sigma \atop \rho\ \beta}\right\} g^{\beta\rho} - \left\{{\rho \atop \rho\ \beta}\right\} g^{\sigma\beta} + \left\{{\beta \atop \rho\ \beta}\right\} g^{\sigma\rho}\right]\sqrt{-g}\end{aligned}$$

This expression collapses by rearrangement and cancellation to

$$\left\{{\alpha \atop \sigma\ \rho}\right\} (\sqrt{-g}\,g^{\sigma\rho})_{|\alpha} - \left\{{\alpha \atop \sigma\ \alpha}\right\} (\sqrt{-g}\,g^{\sigma\rho})_{|\rho} = 2\mathfrak{A} \tag{11.39}$$

with

$$\mathfrak{A} = g^{\rho\sigma}\left[\left\{{\alpha \atop \sigma\ \rho}\right\}\left\{{\beta \atop \alpha\ \beta}\right\} - \left\{{\alpha \atop \beta\ \rho}\right\}\left\{{\beta \atop \alpha\ \sigma}\right\}\right]\sqrt{-g} \tag{11.40}$$

Let us substitute the expression (11.39) into (11.36) and insert (11.36) back into the definition (11.19) of the Riemann scalar density

$$\mathfrak{R} = R\sqrt{-g}$$

We thereby obtain the remarkable identity

$$\begin{aligned}(11.41)\quad \mathfrak{R} = R\sqrt{-g} &= \left(\sqrt{-g}\,g^{\sigma\rho}\left\{{\alpha \atop \sigma\ \alpha}\right\}\right)_{|\rho} \\ &\quad - \left(\sqrt{-g}\,g^{\sigma\rho}\left\{{\alpha \atop \sigma\ \rho}\right\}\right)_{|\alpha} + \mathfrak{A}\end{aligned}$$

The desired decomposition of $\mathfrak{R}$ into an ordinary divergence term and a simple expression involving only the $g_{\mu\nu}$ and their first derivatives is achieved.

Using Gauss's theorem for integrals of ordinary divergences, we now have

$$J = \int_{D^4} \mathfrak{R}\, d^4x = \int_{D^4} \mathfrak{A}\, d^4x + \text{(surface terms)} \tag{11.42}$$

Let us then give a name to the first term of (11.42):

$$H = \int_{D^4} \mathfrak{A}\, d^4x \tag{11.43}$$

Since $\mathfrak{A}$ is *not* a scalar density and depends on the choice of the coordinate system in a complicated manner, the integral H changes with different reference systems. However, given a specific coordinate system and a metric tensor $g_{\mu\nu}$ in it, we can assert that the values of J and H will undergo the same variation if we vary the metric tensor $g_{\mu\nu}$ in the coordinate region D^4 in an arbitrary way, but so that the variation of the $g_{\mu\nu}$ and their derivatives vanish on the boundary of D^4. Thus J and H have the same functional derivatives with respect to the metric tensor $g_{\mu\nu}$. This simple fact will be of great importance in what follows.

Let us next compute δH; since the integrand $\mathfrak{A}$ of H is a function of only the $g_{\mu\nu}$ and their first derivatives, we have, by the usual Euler-Lagrange method,

$$\begin{aligned}(11.44)\quad \delta H &= \delta\int_{D_4} \mathfrak{A}(g^{\mu\nu}, g^{\mu\nu}{}_{|\lambda})\, d^4x \\ &= \int_{D_4}\left[\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}}\,\delta g^{\mu\nu} + \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\delta g^{\mu\nu}{}_{|\lambda}\right] d^4x \\ &= \int_{D_4}\left[\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}} - \left(\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\right)_{|\lambda}\right]\delta g^{\mu\nu}\, d^4x + \text{(surface term)}\end{aligned}$$

For variations which vanish on the boundary of D_4, the surface term will, of course, be zero; thus, equating δJ in (11.35) and δH in (11.44), we obtain

$$\sqrt{-g}\,G_{\mu\nu} = \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}} - \left(\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\right)_{|\lambda} \tag{11.45}$$

This very convenient and interesting expression for the Einstein tensor density is the net result of our variational analysis.

Our search for an expression $\sqrt{-g}\,t_\mu{}^\nu$ is now near an end. The final manipulation consists in relabeling indices in (11.45) and multiplying by $g^{\tau\beta}{}_{|\mu}$:

$$(11.46)\qquad g^{\tau\beta}{}_{|\mu}(\sqrt{-g}\,G_{\tau\beta}) = \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}}\, g^{\tau\beta}{}_{|\mu} - \left(\frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\lambda}}\right)_{|\lambda} g^{\tau\beta}{}_{|\mu}$$
$$= \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}}\, g^{\tau\beta}{}_{|\mu} - \left(\frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\lambda}}\, g^{\tau\beta}{}_{|\mu}\right)_{|\lambda} + \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\lambda}}\, g^{\tau\beta}{}_{|\lambda|\mu}$$

Observe that $\mathfrak{A}$ is a function of only $g^{\tau\beta}$ and $g^{\tau\beta}{}_{|\mu}$; thus

$$(11.47)\qquad \mathfrak{A}_{|\mu} = \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}}\, g^{\tau\beta}{}_{|\mu} + \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\lambda}}\, g^{\tau\beta}{}_{|\lambda|\mu}$$

and (11.44) becomes

$$(11.48)\qquad g^{\tau\beta}{}_{|\mu}(\sqrt{-g}\,G_{\tau\beta}) = \mathfrak{A}_{|\mu} - \left(\frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\lambda}}\, g^{\tau\beta}{}_{|\mu}\right)_{|\lambda}$$
$$= \left(\mathfrak{A}g_\mu{}^\lambda - \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\lambda}}\, g^{\tau\beta}{}_{|\mu}\right)_{|\lambda}$$

By noting that $g^{\tau\beta}g_{\tau\lambda}$ is the invariant Kronecker delta, we see that

$$(11.49)\qquad (g^{\tau\beta}g_{\tau\lambda})_{|\mu} = g^{\tau\beta}{}_{|\mu}g_{\tau\lambda} + g^{\tau\beta}g_{\tau\lambda|\mu} = 0$$

Thus

$$(11.50)\qquad g^{\tau\beta}{}_{|\mu}G_{\tau\beta} = g^{\tau\beta}{}_{|\mu}g_{\tau\lambda}g_{\beta\sigma}G^{\lambda\sigma}$$
$$= -g^{\tau\beta}g_{\tau\lambda|\mu}g_{\beta\sigma}G^{\lambda\sigma} = -g_{\tau\beta|\mu}G^{\tau\beta}$$

so we can reverse the position of the indices on the left side of (11.48) if we reverse the sign:

$$(11.51)\qquad \sqrt{-g}\,g_{\tau\beta|\mu}G^{\tau\beta} = -\left(\mathfrak{A}g_\mu{}^\lambda - \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\lambda}}\, g^{\tau\beta}{}_{|\mu}\right)_{|\lambda}$$

This important result marks the end of our search, for if we define $\sqrt{-g}\,t_\mu{}^\nu$ to be

$$(11.52)\qquad \sqrt{-g}\,t_\mu{}^\nu = \frac{1}{2C}\left(\mathfrak{A}g_\mu{}^\nu - \frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\nu}}\, g^{\tau\beta}{}_{|\mu}\right)$$

then, by using (11.51), we can write (11.18) as

$$(11.53)\qquad (\sqrt{-g}\,T_\mu{}^\nu)_{|\nu} + (\sqrt{-g}\,t_\mu{}^\nu)_{|\nu} = (\sqrt{-g}\,[T_\mu{}^\nu + t_\mu{}^\nu])_{|\nu} = 0$$

Thus the quantity

$$(11.54)\qquad \sqrt{-g}\,(T_\mu{}^\nu + t_\mu{}^\nu) = \sqrt{-g}\,T_\mu{}^\nu + \frac{1}{2C}\,\mathfrak{A}g_\mu{}^\nu - \frac{1}{2C}\,\frac{\partial\mathfrak{A}}{\partial g^{\tau\beta}{}_{|\nu}}\, g^{\tau\beta}{}_{|\mu}$$

has a zero ordinary divergence and represents the density of some conserved quantity. The expression $\sqrt{-g}\,t_\mu{}^\nu$ is usually referred to as the pseudo-tensor of the gravitational field.

The zero divergence of $\sqrt{-g}\,(T_\mu{}^\nu + t_\mu{}^\nu)$ gives rise to a conserved integral quantity in precisely the same way that the zero divergence of $T_\mu{}^\nu$ in flat space gives rise to the energy-momentum four-vector P_μ. Indeed, let us assume that the energy-momentum tensor is different from zero only in a finite part of space and that we can introduce a metric $g_{\mu\nu}$ which tends to the Lorentzian form as we approach infinity. It is evident that the quantity $t_\mu{}^\nu$ will tend to zero if we approach infinity. Let us now integrate (11.53) over all space-time between the hypersurfaces $x^0 = ct_i$ and $x^0 = ct_f$. As in the beginning of this section, we can apply integration by parts to obtain

$$(11.55)\qquad \int_{t_f} \sqrt{-g}\,[T_\mu{}^0 + t_\mu{}^0]\,d^3x - \int_{t_i} \sqrt{-g}\,[T_\mu{}^0 + t_\mu{}^0]\,d^3x = 0$$

in analogy with (11.12). The quantity

$$(11.56)\qquad P_\mu = \int_t \sqrt{-g}\,[T_\mu{}^0 + t_\mu{}^0]\,d^3x$$

is therefore conserved, and may be viewed as the *general relativistic generalization of the energy-momentum four-vector* of special relativity theory.

One must note carefully, however, that the quantity $\sqrt{-g}\,(T_\mu{}^\nu + t_\mu{}^\nu)$ is not a tensor and P_μ is not a generally covariant four-vector. This comes about, of course, because $\mathfrak{A}$ is not a scalar density. Furthermore, it should be noted that we assume that the coordinate system we are using when we integrate (11.53) over D^4 is Lorentzian at the spatial infinity of each coordinate; if this were not true, we could not obtain (11.55). For instance, *in polar coordinates*, (11.55) does *not* follow from (11.53).

Lastly, let us mention that our expression for $t_\mu{}^\nu$ is not unique; many authors use quite different expressions from ours and of course obtain results consistent with these alternative expressions.

It would be tempting to label $t_\mu{}^\nu$ as the energy-momentum tensor of the gravitational field; the remarks of the preceding paragraph, however, indicate that it is not possible to do this in a coordinate-invariant manner. It appears that the intimate connection between geometry, the gravitational field, and the notion of density makes the idea of the energy-momentum density of the gravitational field intrinsically noncovariant.

At this point we can summarize the difficulty of the situation as follows: Since $t_\mu{}^\nu$ is zero in a Lorentzian metric, we may say that we evaluate the gravitational energy-momentum quantity by its deviation from Lorentzian character. We know, however, that we may always introduce locally geodesic coordinates, which, at any chosen point, cause the non-Euclidean character of the geometry to disappear and lead to zero gravitational energy–momentum at that point. Thus the choice of some specific coordinate system means a particular localization of gravitational energy and momentum. It is satisfactory that the energy-momentum balance always comes out in the same way in the large, but the local distribution is coordinate-dependent and cannot be described in an intrinsic coordinate-independent way.

11.3 An Alternative Approach to the Conservation Laws: Energy-Momentum of the Schwarzschild Field

In the previous section we obtained the gravitational-field pseudo-tensor $t_\mu{}^\nu$ by considering a variation of the metric-tensor field. It is also possible to obtain an expression for $t_\mu{}^\nu$ by considering a variation of the coordinate system instead of the metric field, and in some respects the analysis is simpler. This is what we shall do in this section.

Let us again consider the noninvariant quantity

$$H = \int \mathfrak{A}\, d^4x \tag{11.57}$$
$$\mathfrak{A} = \sqrt{-g}\, g^{\sigma\rho}\left[\begin{Bmatrix}\alpha\\ \sigma\ \rho\end{Bmatrix}\begin{Bmatrix}\beta\\ \alpha\ \beta\end{Bmatrix} - \begin{Bmatrix}\alpha\\ \beta\ \rho\end{Bmatrix}\begin{Bmatrix}\beta\\ \alpha\ \sigma\end{Bmatrix}\right]$$

Clearly, $\mathfrak{A}$ involves only the $g^{\mu\nu}$ and their first derivatives, which fact we made use of in the preceding section. For any change of the metric field, the variation of $\mathfrak{A}$ will be

$$\delta\mathfrak{A} = \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}}\,\delta g^{\mu\nu} + \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\delta g^{\mu\nu}{}_{|\lambda} \tag{11.58}$$

Among all possible variations of $g^{\mu\nu}$, we shall investigate those which are due to a small change of coordinates of the form

$$\bar{x}^\alpha = x^\alpha + \epsilon\xi^\alpha \tag{11.59}$$

where ξ^α is an arbitrary function of the x^α, and ϵ is a small constant. This coordinate transformation leads to a new metric tensor

$$\bar{g}^{\mu\nu} = \frac{\partial\bar{x}^\mu}{\partial x^\alpha}\frac{\partial\bar{x}^\nu}{\partial x^\beta}\, g^{\alpha\beta} = \left(\delta^\mu{}_\alpha + \epsilon\frac{\partial\xi^\mu}{\partial x^\alpha}\right)\left(\delta^\nu{}_\beta + \epsilon\frac{\partial\xi^\nu}{\partial x^\beta}\right) g^{\alpha\beta} \tag{11.60}$$
$$= g^{\mu\nu} + \epsilon\left(\frac{\partial\xi^\mu}{\partial x^\alpha}\, g^{\alpha\nu} + \frac{\partial\xi^\nu}{\partial x^\beta}\, g^{\mu\beta}\right) + O(\epsilon^2)$$

Thus the first-order variation of $g^{\mu\nu}$ is

$$\delta g^{\mu\nu} = \epsilon\left(\frac{\partial\xi^\mu}{\partial x^\alpha}\, g^{\alpha\nu} + \frac{\partial\xi^\nu}{\partial x^\beta}\, g^{\mu\beta}\right) = \epsilon(\xi^\mu{}_{|\alpha} g^{\alpha\nu} + \xi^\nu{}_{|\beta} g^{\mu\beta}) \tag{11.61}$$

Some care must be exercised in calculating $\delta g^{\mu\nu}{}_{|\lambda}$; by definition,

$$\bar{g}^{\mu\nu}{}_{|\lambda} = \frac{\partial\bar{g}^{\mu\nu}}{\partial\bar{x}^\lambda} = \frac{\partial\bar{g}^{\mu\nu}}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial\bar{x}^\lambda} \tag{11.62}$$

From the transformation equation (11.59), we have

$$\frac{\partial x^\alpha}{\partial\bar{x}^\lambda} = \delta^\alpha{}_\lambda - \epsilon\frac{\partial\xi^\alpha}{\partial x^\lambda} + O(\epsilon^2) \tag{11.63}$$

Substituting (11.60) and (11.63) in (11.62), we obtain

$$\frac{\partial\bar{g}^{\mu\nu}}{\partial\bar{x}^\lambda} = \frac{\partial}{\partial x^\alpha}\left[g^{\mu\nu} + \epsilon\left(\frac{\partial\xi^\mu}{\partial x^\tau}\, g^{\tau\nu} + \frac{\partial\xi^\nu}{\partial x^\beta}\, g^{\mu\beta}\right)\right]\left[\delta^\alpha{}_\lambda - \epsilon\frac{\partial\xi^\alpha}{\partial x^\lambda}\right] + O(\epsilon^2) \tag{11.64}$$
$$= \frac{\partial g^{\mu\nu}}{\partial x^\lambda} + \epsilon\left[\frac{\partial}{\partial x^\lambda}\left(\frac{\partial\xi^\mu}{\partial x^\tau}\, g^{\tau\nu} + \frac{\partial\xi^\nu}{\partial x^\beta}\, g^{\mu\beta}\right) - \frac{\partial g^{\mu\nu}}{\partial x^\alpha}\frac{\partial\xi^\alpha}{\partial x^\lambda}\right] + O(\epsilon^2)$$

Hence the variation of $g^{\mu\nu}{}_{|\lambda}$ is

$$\delta g^{\mu\nu}{}_{|\lambda} = \epsilon[(\xi^\mu{}_{|\tau} g^{\tau\nu} + \xi^\nu{}_{|\beta} g^{\mu\beta})_{|\lambda} - g^{\mu\nu}{}_{|\alpha}\xi^\alpha{}_{|\lambda}] \tag{11.65}$$
$$= \epsilon[g^{\tau\nu}{}_{|\lambda}\xi^\mu{}_{|\tau} + g^{\mu\beta}{}_{|\lambda}\xi^\nu{}_{|\beta} - g^{\mu\nu}{}_{|\alpha}\xi^\alpha{}_{|\lambda} + \xi^\mu{}_{|\tau|\lambda} g^{\tau\nu} + \xi^\nu{}_{|\beta|\lambda} g^{\mu\beta}]$$

The variation of $\sqrt{-g}$ is easily obtained using the general relation (11.30) and (11.61):

$$\delta\sqrt{-g} = \frac{-\sqrt{-g}}{2}\, g_{\mu\nu}\,\delta g^{\mu\nu} \tag{11.66}$$
$$= -\frac{\epsilon}{2}\sqrt{-g}\, g_{\mu\nu}(\xi^\mu{}_{|\alpha} g^{\alpha\nu} + \xi^\nu{}_{|\beta} g^{\mu\beta}) = -\epsilon\sqrt{-g}\,\xi^\alpha{}_{|\alpha}$$

The variations in $g^{\mu\nu}$, $g^{\mu\nu}{}_{|\lambda}$, and $\sqrt{-g}$ which we have calculated in (11.61), (11.65), and (11.66) are general and hold for any coordinate transformation (11.59). We wish to use these relations to calculate $\delta\mathfrak{A}$; to do this we shall first use them to establish a very interesting relation involving $\mathfrak{A}$, which will in turn allow us to obtain a very simple and elegant form for $\delta\mathfrak{A}$. Let us specify the ξ^α to be linear functions of the coordinates x^μ. Equation (11.59) is then a linear transformation of coordinates; under such a linear transformation the Christoffel symbols transform as tensors, as is evident from (2.5). Thus the expression $\mathfrak{A}$ is a scalar density under the restricted class of linear transformations, which is apparent from (11.57). Since $\mathfrak{A}$ is a scalar density under (11.59), it has a very simple variation, which is, in fact, due entirely to the variation of the factor $\sqrt{-g}$ under the transformation (11.59). Indeed, we see that

$$\delta\mathfrak{A} = \left(\frac{\mathfrak{A}}{\sqrt{-g}}\right)\delta\sqrt{-g} \tag{11.67}$$

since $\mathfrak{A}/\sqrt{-g}$ is a scalar invariant. From (11.66) we therefore have

$$\delta\mathfrak{A} = -\epsilon\xi^\alpha{}_{|\alpha}\mathfrak{A} \tag{11.68}$$

for a linear transformation (11.59).

On the other hand, we may also calculate the variation $\delta\mathfrak{A}$ for a linear transformation (11.59) by substituting the variations (11.61) and (11.65) into the general expression for $\delta\mathfrak{A}$ in (11.58); since in the present case ξ^α is a linear function and has zero second derivatives, we obtain

$$\begin{aligned}\delta\mathfrak{A} &= \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}}\,\delta g^{\mu\nu} + \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\delta g^{\mu\nu}{}_{|\lambda} \\ &= 2\epsilon\left(\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}}\,\xi^\mu{}_{|\alpha}g^{\alpha\nu} + \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\xi^\mu{}_{|\alpha}g^{\alpha\nu}{}_{|\lambda}\right) - \epsilon\,\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,g^{\mu\nu}{}_{|\alpha}\xi^\alpha{}_{|\lambda}\end{aligned} \tag{11.69}$$

If we now compare (11.68) and (11.69) and note that the $\xi^\alpha{}_{|\beta}$ are arbitrary constants, we arrive at a remarkable differential identity for $\mathfrak{A}$:

$$\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}}\,g^{\alpha\nu} + \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,g^{\alpha\nu}{}_{|\lambda} - \frac{1}{2}\frac{\partial\mathfrak{A}}{\partial g^{\beta\nu}{}_{|\alpha}}\,g^{\beta\nu}{}_{|\mu} = -\tfrac{1}{2}\mathfrak{A}g_\mu{}^\alpha \tag{11.70}$$

It should be carefully noted that this identity was obtained by using variational methods and a linear coordinate transformation, but the identity is quite general and is completely independent of any coordinate transformation used in its derivation.

It is clear that, by substituting the variations of $g^{\mu\nu}$ and $g^{\mu\nu}{}_{|\lambda}$ in (11.61) and (11.65) into the variation of $\mathfrak{A}$ in (11.58), we can compute $\delta\mathfrak{A}$ for an arbitrary ξ^α and indeed could have computed $\delta\mathfrak{A}$ without bothering to obtain the identity (11.70); we shall now see, however, that (11.70) greatly simplifies the form of $\delta\mathfrak{A}$ and is well worth the price of its derivation. The substitutions into (11.58) give

$$\begin{aligned}\delta\mathfrak{A} &= 2\epsilon\xi^\mu{}_{|\alpha}\left(\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}}\,g^{\alpha\nu} + \frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,g^{\alpha\nu}{}_{|\lambda} - \frac{1}{2}\frac{\partial\mathfrak{A}}{\partial g^{\beta\nu}{}_{|\alpha}}\,g^{\beta\nu}{}_{|\mu}\right) \\ &\quad + 2\epsilon\,\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\xi^\mu{}_{|\tau|\lambda}g^{\tau\nu}\end{aligned} \tag{11.71}$$

The straightforward substitution of (11.70) then gives the much simpler result

$$\delta\mathfrak{A} = -\epsilon\mathfrak{A}\xi^\alpha{}_{|\alpha} + 2\epsilon\,\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\xi^\mu{}_{|\tau|\lambda}g^{\tau\nu} \tag{11.72}$$

which is quite general and holds for any coordinate variation (11.59).

Having obtained these formal results, let us return to physics. For a variation of the metric-tensor field which vanishes on the boundary of D^4, we found in the preceding section that $\delta J = \delta H$; J and H are defined in (11.20) and (11.43). If we choose the ξ^α in (11.59) to be zero and have zero first and second derivatives on the boundary of D^4, it is clear from (11.59) that $\delta g^{\mu\nu} = \delta g^{\mu\nu}{}_{|\lambda} = 0$ on the boundary of D^4, so $\delta H = \delta J$. However, we know that J is a scalar invariant, and thus must have a zero variation under any coordinate variation; thus, for this special coordinate variation,

$$\delta J = \delta H = \delta\int_{D^4}\mathfrak{A}\,d^4x = 0 \tag{11.73}$$

We may write this more conveniently by taking note of the fact that $\sqrt{-g}\,d^4x$ is a scalar invariant and has zero variation under a coordinate variation; then, since the range of the old and the new variables is the same, namely D^4, we have

$$\delta H = \delta\int_{D^4}\left(\frac{\mathfrak{A}}{\sqrt{-g}}\right)\sqrt{-g}\,d^4x = \int_{D^4}\delta\left(\frac{\mathfrak{A}}{\sqrt{-g}}\right)\sqrt{-g}\,d^4x = 0 \tag{11.74}$$

The variation of $\mathfrak{A}/\sqrt{-g}$ is easily obtained from the variation of $\mathfrak{A}$ in (11.72) and the variation of $\sqrt{-g}$ in (11.66):

$$\begin{aligned}\sqrt{-g}\,\delta\left(\frac{\mathfrak{A}}{\sqrt{-g}}\right) &= \delta\mathfrak{A} + \sqrt{-g}\,\mathfrak{A}\delta\left(\frac{1}{\sqrt{-g}}\right) \\ &= \delta\mathfrak{A} - \frac{\mathfrak{A}}{\sqrt{-g}}\,\delta\sqrt{-g} = 2\epsilon\,\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\xi^\mu{}_{|\tau|\lambda}g^{\tau\nu}\end{aligned} \tag{11.75}$$

Thus the null variation of H implies

$$\delta H = 2\epsilon\int_{D^4}\frac{\partial\mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}}\,\xi^\mu{}_{|\tau|\lambda}g^{\tau\nu}\,d^4x = 0 \tag{11.76}$$

Since ξ^μ and $\xi^\mu{}_{|\tau}$ vanish on the boundary of D^4, we may integrate by parts twice to obtain an equivalent expression

$$\delta H = 2\epsilon \int_{D^4} \left(\frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}} g^{\tau\nu} \right)_{|\tau|\lambda} \xi^\mu \, d^4x = 0 \tag{11.77}$$

Since ξ^μ is arbitrary inside D^4, we therefore obtain, finally, the divergence law

$$\left(\frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}} g^{\tau\nu} \right)_{|\tau|\lambda} = 0 \tag{11.78}$$

It is thus apparent that the expression

$$\sqrt{-g}\, F_\mu{}^\tau = \left(\frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}} g^{\tau\nu} \right)_{|\lambda} \tag{11.79}$$

is associated with some conserved physical quantity.

Let us work out $F_\mu{}^\tau$ explicitly to see what sort of physical quantity it represents:

$$\sqrt{-g}\, F_\mu{}^\tau = \left(\frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}} \right)_{|\lambda} g^{\tau\nu} + \frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}} g^{\tau\nu}{}_{|\lambda} \tag{11.80}$$

By solving the useful identity (11.70) for $(\partial \mathfrak{A}/\partial g^{\mu\nu})_{|\lambda} g^{\tau\nu}{}_{|\lambda}$ and substituting the result into (11.80), we arrive at

$$\sqrt{-g}\, F_\mu{}^\tau = \left[\left(\frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}} \right)_{|\lambda} - \frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}} \right] g^{\tau\nu} + \frac{1}{2} \frac{\partial \mathfrak{A}}{\partial g^{\beta\nu}{}_{|\tau}} g^{\beta\nu}{}_{|\mu} - \tfrac{1}{2} \mathfrak{A} g_\mu{}^\tau \tag{11.81}$$

Finally, let us invoke Eq. (11.45), which says that the first bracket in (11.81) is precisely $-\sqrt{-g}\, G_{\mu\nu}$; this gives

$$\sqrt{-g}\, F_\mu{}^\tau = -\sqrt{-g}\, G_{\mu\nu} - \tfrac{1}{2} \mathfrak{A} g_\mu{}^\tau + \frac{1}{2} \frac{\partial \mathfrak{A}}{\partial g^{\beta\nu}{}_{|\tau}} g^{\beta\nu}{}_{|\mu} \tag{11.82}$$

or, by using the field equations (11.9),

$$\sqrt{-g}\, F_\mu{}^\tau = -C \sqrt{-g}\, T_\mu{}^\tau - \tfrac{1}{2} \mathfrak{A} g_\mu{}^\tau + \frac{1}{2} \frac{\partial \mathfrak{A}}{\partial g^{\beta\nu}{}_{|\tau}} g^{\beta\nu}{}_{|\mu} \tag{11.83}$$

This last form we immediately recognize as being essentially the same conserved quantity as we obtained in the preceding section; indeed, by comparing with (11.54), we see that

$$F_\mu{}^\tau = -C[T_\mu{}^\tau + t_\mu{}^\tau] \tag{11.84}$$

Our alternative method to find the pseudo-tensor $F_\mu{}^\tau$ with zero divergence is of great significance for the general mathematical approach to field equations in general relativity. It should be observed that we carried out many operations without using the explicit dependence of $\mathfrak{A}$ on the metric tensor $g_{\mu\nu}$. In fact, suppose we had started with an arbitrary expression $\mathfrak{A} = A \sqrt{-g}$, which depends only on the $g_{\mu\nu}$ and their first derivatives and on other field quantities that are independent of the metric tensor. To simplify matters, let us even assume that A is an actual scalar, i.e., is invariant under any change of variables. All derivations from (11.58) to (11.81) would have remained valid. Only to pass from (11.81) to (11.82) did we use the identity (11.45), which is based on the fundamental variational formula (11.35). But formula (11.79) is already sufficient to construct from the scalar density $\mathfrak{A}$ a pseudo-tensor density $F_\mu{}^\tau \sqrt{-g}$ whose ordinary divergence is zero. Such a pseudo-tensor density can always be interpreted as connected with a conserved quantity.

section and discuss the identity (11.84). Although we obtain the same conserved quantity by varying the coordinates as we obtained previously by varying the metric-tensor field, the present approach has the virtue that the conserved $F_\mu{}^\tau$ itself appears as an ordinary divergence in (11.79). This fact allows a simplification of the conservation law in special cases. Indeed, by the general procedure of the last section, it is evident that

$$P_\mu = \frac{-1}{C} \int_{V^3} \left(\frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|\lambda}} g^{0\nu} \right)_{|\lambda} d^3x \qquad C = -\frac{8\pi\kappa}{c^2} \tag{11.85}$$

is a conserved quantity if V^3 is the entire space at a given time. In the special case that the metric is independent of time, we need not sum over $\lambda = 0$, so we can use Gauss's theorem in three dimensions to obtain the energy-momentum content of any finite region V^3,

$$P_\mu = \frac{-1}{C} \int_{S^2} \frac{\partial \mathfrak{A}}{\partial g^{\mu\nu}{}_{|j}} g^{0\nu} n_j \, dS \tag{11.86}$$

where n_j is a unit normal to the surface element of S^2, which is the boundary of V^3. We therefore have the interesting and curious result that the generalized energy momentum P_μ of a volume V^3 may be determined from the values of the metric-tensor field and its derivatives *on the surface of* V^3; the detailed behavior of the field inside V^3 is irrelevant.

To conclude this section we shall compute P_0 for a Schwarzschild field using Eq. (11.86). This calculation will serve to illustrate the physics and to clarify the usefulness and the limited validity of the concept of the generalized energy momentum P_μ. As we stressed at the end of the previous section, we must use coordinates for which the metric is Lorentzian at spatial infinity, so the standard form of the Schwarzschild metric in polar coordinates will not do; instead, we shall use the isotropic form of Sec. 6.2, expressed in terms of the coordinates of special relativity:

$$ds^2 = \frac{(1 - m/2\rho)^2}{(1 + m/2\rho)^2} c^2\, dt^2 - \left(1 + \frac{m}{2\rho}\right)^4 (dx^2 + dy^2 + dz^2) \tag{11.87}$$

This form is clearly Lorentzian at infinity, $\rho = \infty$. The metric tensor is therefore

(11.88)
$$g_{\mu\nu} = \begin{pmatrix} \dfrac{(1 - m/2\rho)^2}{(1 + m/2\rho)^2} & 0 & 0 & 0 \\ 0 & -\left(1 + \dfrac{m}{2\rho}\right)^4 & 0 & 0 \\ 0 & 0 & -\left(1 + \dfrac{m}{2\rho}\right)^4 & 0 \\ 0 & 0 & 0 & -\left(1 + \dfrac{m}{2\rho}\right)^4 \end{pmatrix}$$

and its inverse is

(11.89)
$$g^{\mu\nu} = \begin{pmatrix} \dfrac{(1 + m/2\rho)^2}{(1 - m/2\rho)^2} & 0 & 0 & 0 \\ 0 & -\left(1 + \dfrac{m}{2\rho}\right)^{-4} & 0 & 0 \\ 0 & 0 & -\left(1 + \dfrac{m}{2\rho}\right)^{-4} & 0 \\ 0 & 0 & 0 & -\left(1 + \dfrac{m}{2\rho}\right)^{-4} \end{pmatrix}$$

Since the metric tensor $g^{\mu\nu}$ is diagonal, the quantity P_0 which we wish to calculate is, from (11.86),

$$P_0 = \frac{-1}{C} \int_{S^2} \frac{\partial \mathfrak{A}}{\partial g^{00}{}_{|j}}\, g^{00} n_j\, dS \tag{11.90}$$

For convenience we shall take S^2 to be the surface of a sphere of radius $\rho = R$.

The problem now is to compute the quantity $\partial\mathfrak{A}/\partial g^{00}{}_{|j}$, where $\mathfrak{A}$ is considered to be a function of $g^{\mu\nu}$ and its first derivatives. Consider the

first term of

$$\mathfrak{A} = \sqrt{-g}\, g^{\sigma\rho} \left[\left\{ \begin{matrix} \alpha \\ \sigma\ \rho \end{matrix} \right\} \left\{ \begin{matrix} \beta \\ \alpha\ \beta \end{matrix} \right\} - \left\{ \begin{matrix} \alpha \\ \beta\ \sigma \end{matrix} \right\} \left\{ \begin{matrix} \beta \\ \alpha\ \rho \end{matrix} \right\} \right] \tag{11.91}$$

This term may be rewritten, with the use of the definition of the Christoffel symbols and the symmetry of $g^{\mu\nu}$, as

$$\begin{aligned} \sqrt{-g}\, g^{\sigma\rho} \left\{ \begin{matrix} \alpha \\ \sigma\ \rho \end{matrix} \right\} \left\{ \begin{matrix} \beta \\ \alpha\ \beta \end{matrix} \right\} \\ = \frac{\sqrt{-g}}{4} g^{\sigma\rho} g^{\alpha\tau} (g_{\rho\tau|\sigma} + g_{\sigma\tau|\rho} - g_{\sigma\rho|\tau}) g^{\beta\kappa} g_{\beta\kappa|\alpha} \end{aligned} \tag{11.92}$$

In order to express this in terms of the derivatives of $g^{\mu\nu}$ instead of the derivatives of $g_{\mu\nu}$, we make use of the convenient elementary relation $g_{\mu\nu|\lambda} g^{\nu\epsilon} = -g^{\nu\epsilon}{}_{|\lambda} g_{\mu\nu}$ and obtain

$$\begin{aligned} \sqrt{-g}\, g^{\sigma\rho} \left\{ \begin{matrix} \alpha \\ \sigma\ \rho \end{matrix} \right\} \left\{ \begin{matrix} \beta \\ \alpha\ \beta \end{matrix} \right\} \\ = \frac{\sqrt{-g}}{4} g^{\alpha\tau} (g_{\rho\tau} g^{\sigma\rho}{}_{|\sigma} + g_{\sigma\tau} g^{\sigma\rho}{}_{|\rho} - g_{\sigma\rho} g^{\sigma\rho}{}_{|\tau}) g_{\beta\kappa} g^{\beta\kappa}{}_{|\alpha} \end{aligned} \tag{11.93}$$

This is easily differentiated with respect to $g^{00}{}_{|j}$:

$$\begin{aligned} \frac{\partial}{\partial g^{00}{}_{|j}} \left(\sqrt{-g}\, g^{\sigma\rho} \left\{ \begin{matrix} \alpha \\ \sigma\ \rho \end{matrix} \right\} \left\{ \begin{matrix} \beta \\ \alpha\ \beta \end{matrix} \right\} \right) \\ = \frac{\sqrt{-g}}{2} g_{00} g^{j\tau} (g_{\rho\tau} g^{\sigma\rho}{}_{|\sigma} - g_{\sigma\rho} g^{\sigma\rho}{}_{|\tau}) \end{aligned} \tag{11.94}$$

The second term of $\mathfrak{A}$ is also easily handled; note first that

$$\begin{aligned} \frac{\partial}{\partial g^{00}{}_{|j}} \left(\sqrt{-g}\, g^{\sigma\rho} \left\{ \begin{matrix} \alpha \\ \beta\ \sigma \end{matrix} \right\} \left\{ \begin{matrix} \beta \\ \alpha\ \rho \end{matrix} \right\} \right) &= \sqrt{-g}\, g^{\sigma\rho} \left(\left\{ \begin{matrix} \alpha \\ \beta\ \sigma \end{matrix} \right\} \frac{\partial}{\partial g^{00}{}_{|j}} \left\{ \begin{matrix} \beta \\ \alpha\ \rho \end{matrix} \right\} + \left\{ \begin{matrix} \beta \\ \alpha\ \rho \end{matrix} \right\} \frac{\partial}{\partial g^{00}{}_{|j}} \left\{ \begin{matrix} \alpha \\ \beta\ \sigma \end{matrix} \right\} \right) \\ &= 2\sqrt{-g}\, g^{\sigma\rho} \left\{ \begin{matrix} \alpha \\ \beta\ \sigma \end{matrix} \right\} \frac{\partial}{\partial g^{00}{}_{|j}} \left\{ \begin{matrix} \beta \\ \alpha\ \rho \end{matrix} \right\} \end{aligned} \tag{11.95}$$

This considerably simplifies the calculation. Next, using the definition of the Christoffel symbol, we write this as

$$\frac{\partial}{\partial g^{00}{}_{|j}} \left(\sqrt{-g}\, g^{\sigma\rho} \left\{ \begin{matrix} \alpha \\ \beta\ \sigma \end{matrix} \right\} \left\{ \begin{matrix} \beta \\ \alpha\ \rho \end{matrix} \right\} \right) \tag{11.96}$$

$$= \sqrt{-g} \begin{Bmatrix} \alpha \\ \beta \ \sigma \end{Bmatrix} \frac{\partial}{\partial g^{00}{}_{|j}} (-g_{\rho\tau} g^{\sigma\rho} g^{\beta\tau}{}_{|\alpha} - g_{\alpha\tau} g^{\sigma\rho} g^{\beta\tau}{}_{|\rho} + g_{\alpha\rho} g^{\beta\tau} g^{\sigma\rho}{}_{|\tau})$$

which may be immediately differentiated to give

$$\frac{\partial}{\partial g^{00}{}_{|j}} \left(\sqrt{-g}\, g^{\sigma\rho} \begin{Bmatrix} \alpha \\ \beta \ \sigma \end{Bmatrix} \begin{Bmatrix} \beta \\ \alpha \ \rho \end{Bmatrix} \right) = -\sqrt{-g} \begin{Bmatrix} j \\ 0 \ 0 \end{Bmatrix} \tag{11.97}$$

Since the metric is time-independent, this further simplifies to

$$\frac{\partial}{\partial g^{00}{}_{|j}} \left(\sqrt{-g}\, g^{\sigma\rho} \begin{Bmatrix} \alpha \\ \beta \ \sigma \end{Bmatrix} \begin{Bmatrix} \beta \\ \alpha \ \rho \end{Bmatrix} \right) = -\sqrt{-g} \begin{Bmatrix} j \\ 0 \ 0 \end{Bmatrix} \tag{11.98}$$

$$= -\frac{\sqrt{-g}}{2} g^{j\lambda} (g_{\lambda 0|0} + g_{\lambda 0|0} - g_{00|\lambda}) = \frac{\sqrt{-g}}{2} g^{j\lambda} g_{00|\lambda}$$

Combining (11.98) and (11.94), we finally have

$$\frac{\partial \mathfrak{A}}{\partial g^{00}{}_{|j}} = \frac{\sqrt{-g}}{2} [g_{00} g^{\sigma j}{}_{|\sigma} - g_{00} g^{j\tau} g_{\sigma\rho} g^{\sigma\rho}{}_{|\tau} - g^{j\sigma} g_{00|\sigma}] \tag{11.99}$$

This is a general result valid for any time-independent metric.

For the specific metric tensor (11.89) which is diagonal and spatially isotropic, we obtain from (11.99) after an easy rearrangement of terms

$$g^{00} \frac{\partial \mathfrak{A}}{\partial g^{00}{}_{|j}} = -\sqrt{-g}\, g^{11}{}_{|j} \tag{11.100}$$

Using the explicit form (11.89) for $g^{\mu\nu}$, we have, finally,

$$g^{00} \frac{\partial \mathfrak{A}}{\partial g^{00}{}_{|j}} = \frac{2mx^j}{\rho^3} \left(1 - \frac{m}{2\rho}\right) \tag{11.101}$$

This is the exact integrand of (11.90) in explicit form; we need only integrate over a sphere of radius R to find the total "energy" P_0 inside the sphere. We thus have to calculate the integral (11.90) in the form

$$P_0 = \frac{c^2}{8\pi\kappa} \int_{S^2} 2m \frac{1}{\rho^3} x^j n_j \left(1 - \frac{m}{2\rho}\right) dS \tag{11.102}$$

One should now beware of an error which is frequently committed. One should not evaluate the surface element dS in the Schwarzschild metric, but in the Euclidean metric of the marker space of the x^j. Indeed, we

obtained the surface integral (11.86) from the volume integral (11.85) by use of the Gauss integral theorem, and n_j and dS are to be understood in the sense required by classical integral calculus. Hence dS is the ordinary Euclidean surface element of a sphere S^2 with radius $\rho = R$. Clearly, $x^j n_j = R$, and (11.102) reduces to

$$P_0 = \frac{mc^2}{\kappa} \left(1 - \frac{m}{2R}\right) \tag{11.103}$$

Let us replace the mass parameter m by the usual mass M of the source of the Schwarzschild field as given by (6.54). We then find that

$$P_0 = M \left(1 - \frac{\kappa M}{2c^2 R}\right) \tag{11.104}$$

As we might expect, the asymptotic value of P_0 is M, the total mass of the particle or spherical body, in agreement with special relativity theory. For finite values of R, this formula would give us the gravitational energy content within a sphere of radius R. Interestingly enough, the energy of gravitation starts with the value zero for $R = m/2$; that is, the entire mass is due to the contributions of the gravitational field outside of the radius $R = m/2$ in the isotropic Schwarzschild line element. From this critical radius the energy content increases with the radius of the sphere. At the distance R from the center of the field we have an energy density

$$\Delta P_0 = \frac{\kappa M^2}{8\pi R^4 c^2} \tag{11.105}$$

This result stands in complete analogy to the classical formula for the electrical energy density around a charged singularity with total charge e. In our units where energy is measured as mass, we have, for the corresponding energy density,

$$\Delta P_0 = \frac{e^2}{8\pi R^4 c^2} \tag{11.106}$$

The absence of a κ constant in the electrical energy density term is due to the choice of the units for the electrical charge such that no constant κ occurs in the Coulomb law. However, the same distance dependence in Newton's and Coulomb's laws of attraction leads to the same radial dependence of the energy density.

11.4 Variational Principles in General Relativity Theory: A Lagrangian Density for the Gravitational Field

The formal considerations of the preceding sections are relevant to a variational formulation of various physical theories within the framework of general relativity. In this section we shall first show that the equations of the gravitational field in empty space can easily be expressed in a variational form: The Lagrangian of this variational principle has in fact been obtained in the preceding sections. Then we shall show how the Lagrangian for the gravitational field in empty space can be extended in such a way that the resultant variational principle will describe also the effect of electromagnetic fields. Our considerations will be illustrative of the great flexibility and power of the variational method and will suggest the possibility of numerous applications in various fields of physics.

We begin with Eqs. (11.20) and (11.35) of Sec. 11.2, which together imply

$$\delta J = \delta \int_{D^4} R\sqrt{-g}\, d^4x = \int_{D^4} G_{\mu\nu}\sqrt{-g}\, \delta g^{\mu\nu}\, d^4x \tag{11.107}$$

Since the equations of the gravitational field in empty space are $G_{\mu\nu} = 0$, we see that we may express the 10 field equations by the single variational principle

$$\delta J = \delta \int_{D^4} R\sqrt{-g}\, d^4x = 0 \tag{11.108}$$

The variations admitted in (11.108) are such that the metric tensor and its first derivatives do not vary on the boundary of D^4.

We may interpret (11.107) to mean that the Einstein tensor density $G_{\mu\nu}\sqrt{-g}$ is the variational derivative of the scalar density $R\sqrt{-g}$ under a variation of the metric tensor. Furthermore, we see that $R\sqrt{-g}$ can be interpreted as the Lagrangian density of the gravitational field in empty space. In Sec. 10.4 we have shown that the field equations of gravitation in nonempty space may be written as in (10.80)

$$G_{\mu\nu} = CT_{\mu\nu} \qquad C = -\frac{8\pi\kappa}{c^2} \tag{11.109}$$

The appearance of the Einstein tensor in these more general equations and our above results suggest the possibility of extending the variational formulation (11.108) to nonempty space.

Consider a physical system characterized by a specific energy-momentum tensor $T_{\mu\nu}$. Our aim is then to construct a scalar density $L\sqrt{-g}$, which depends on the metric tensor $g^{\mu\nu}$ and possibly other field variables (electromagnetic potentials, velocity fields, etc.) such that

$$\delta \int_{D^4} L\sqrt{-g}\, d^4x = \int_{D^4} T_{\mu\nu}\sqrt{-g}\, \delta g^{\mu\nu}\, d^4x \tag{11.110}$$

under a variation of the metric field as considered above. If we could do this, the gravitational field equations in nonempty space (11.109) could be expressed in the variational form

$$\delta \int_{D^4} (R - CL)\sqrt{-g}\, d^4x = 0 \tag{11.111}$$

where only the metric-tensor field is varied. The scalar density $(R - CL)\sqrt{-g}$ would represent the Lagrangian density of the extended system.

Observe, however, that Eqs. (11.109), or equivalently (11.111), do not completely describe the physical system, but only its gravitational aspects. It is often possible to choose the expression L in such a way that the variation of the same integral as in (11.111) with respect to the additional field variables associated with $T_{\mu\nu}$ vanishes as a consequence of the additional field equations for these variables. In this case the unrestricted variation principle (11.111) would give a complete description of the physical system. We should then consider $(R - CL)\sqrt{-g}$ as the Lagrangian density of the entire system and $R\sqrt{-g}$ as the contribution of the gravitational field to the total Lagrangian.

We illustrate these general remarks by an example which is of considerable importance in its own right, the electromagnetic field. Let $F_{\mu\nu}$ be the antisymmetric tensor which describes the electromagnetic field (as discussed in Secs. 4.1 and 9.1) and define the scalar

$$L = AF_{\mu\nu}F^{\mu\nu} = Ag^{\mu\alpha}g^{\nu\beta}F_{\mu\nu}F_{\alpha\beta} \tag{11.112}$$

with a constant A, which will be conveniently prescribed later. In this case we can easily calculate the variation of the integral (11.110) by the use of (11.30) and find, considering $F_{\mu\nu}$ independent of $g_{\mu\nu}$,

$$\delta \int_{D^4} L\sqrt{-g}\, d^4x = -2A \int_{D^4} [F_\mu{}^\rho F_{\rho\nu} + \tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta}g_{\mu\nu}]\sqrt{-g}\, \delta g^{\mu\nu}\, d^4x \tag{11.113}$$

On the other hand, we showed in (10.69) that the energy-momentum tensor of the electromagnetic field in empty space has the form

$$T_{\mu\nu} = F_{\mu}{}^{\rho}F_{\rho\nu} + \tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta}g_{\mu\nu} \tag{11.114}$$

Comparison of (11.113) and (11.114) suggests the choice $A = -\frac{1}{2}$, in which case (11.113) agrees precisely with (11.110). Thus one possible form of the Lagrangian density of the combined gravitational and electromagnetic field in matter-free space has been established to be

$$(R - CL)\sqrt{-g} = \left(R + \frac{C}{2}F_{\alpha\beta}F^{\alpha\beta}\right)\sqrt{-g} \tag{11.115}$$

Surprisingly enough, the same Lagrangian density (11.115) which leads to the Einstein field equations of gravitation if we vary the metric tensor leads also to the Maxwell field equations of electromagnetism if we vary the potentials of the electromagnetic field. To demonstrate this statement, we remind the reader that one set of Maxwell equations, namely,

$$\{F_{\mu\nu|\lambda}\} = 0 \tag{11.116}$$

expresses the fact that $F_{\mu\nu}$ is a closed tensor (Sec. 4.1) and that it possesses a vector potential ϕ_μ. That is, $F_{\mu\nu}$ may be written in the form

$$F_{\mu\nu} = \phi_{\mu|\nu} - \phi_{\nu|\mu} \tag{11.117}$$

Conversely, if we introduce an arbitrary four-vector ϕ_μ and *define* $F_{\mu\nu}$ by (11.117), the set of equations (11.116) will be automatically fulfilled. We consider, therefore, the vector potential ϕ_μ as the independent field variable which must obey only the remaining set of Maxwell equations (4.63) in empty space,

$$F^{\mu\nu}{}_{\|\nu} = \frac{1}{\sqrt{-g}}(F^{\mu\nu}\sqrt{-g})_{|\nu} = 0 \tag{11.118}$$

We shall show now that these equations are precisely the Euler-Lagrange equations of the variational problem (11.111) for the Lagrangian (11.115) under a variation of the potential ϕ_μ. Indeed, observe that $R\sqrt{-g}$ is independent of the vector potential ϕ_μ; so from (11.115) we obtain the Euler-Lagrange equations

$$-\left[\frac{\partial(L\sqrt{-g})}{\partial\phi_{\mu|\nu}}\right]_{|\nu} = 2C(F^{\mu\nu}\sqrt{-g})_{|\nu} = 0 \tag{11.119}$$

which are identical with (11.118) as asserted.

In summary, we have shown that the combined Einstein and Maxwell field equations in matter-free space can be condensed into the single

variational principle

$$\delta\int_{D^4}\left(R + \frac{C}{2}F_{\alpha\beta}F^{\alpha\beta}\right)\sqrt{-g}\,d^4x = 0 \tag{10.135}$$

We obtain Einstein's equations if we vary the metric potentials $g_{\mu\nu}$, and Maxwell's equations for empty space if we vary the electromagnetic potentials ϕ_μ. The variational condition (11.120) summarizes all differential equations of the theory at points of space-time where no particles are located.

Lastly, let us remark that it is possible to extend the above variational principle to include systems with charges and masses. It is well known from special relativity that the Maxwell equations can be derived from a variational principle with a Lagrangian proportional to

$$L = \tfrac{1}{2}F_{\mu\nu}F^{\mu\nu} + 2\phi_\mu s^\mu \tag{11.121}$$

Here s^μ is the current four-vector (Sec. 4.1). The ϕ_μ are considered as independent variables and $F_{\mu\nu}$ is defined by (11.117). It is immediate that the variational problem

$$\delta\int_{D^4} L\sqrt{-g}\,d^4x = 0 \tag{11.122}$$

leads to the Euler-Lagrange equations

$$\left[\frac{\partial(L\sqrt{-g})}{\partial\phi_{\mu|\nu}}\right]_{|\nu} - \frac{\partial(L\sqrt{-g})}{\partial\phi_\mu} = (F^{\mu\nu}\sqrt{-g})_{|\nu} - s^\mu\sqrt{-g} = 0 \tag{11.123}$$

which are identical with the Maxwell equations (4.63) even in the case of general relativity theory. However, the contribution of the masses and their velocities which must be associated with the charges requires additional terms in the complete Lagrangian. We shall not enter into the laborious discussion of the entire system of gravitational fields, electromagnetic fields, charges, and masses.

In this section we have shown that a combined theory of the gravitational and electromagnetic fields in matter-free space can be expressed in very elegant form. However, we wish to emphasize that this is only a demonstration of the power of the formal mathematical tools and by no means a step in the direction of a unified field theory. The aim of a unified field theory is to imbed electromagnetic phenomena into geometry in a way analogous to that done for gravitation. The discussion of this problem will be taken up in Chap. 15.

11.5 The Scalar Tensor Variation of Relativity Theory

We may use the mathematical developments of the preceding sections to study an interesting variant of general relativity theory. The problem of understanding the numerical value of a given physical constant has always intrigued physicists. Indeed one may consider the reduction of the number of arbitrary or unrelated physical constants to be a measure of the overall progress of physics. The puzzling aspect of arbitrary physical constants is especially evident in the case of the gravitational constant, $\kappa = 6.67 \times 10^{-8}$ dyne-cm/g^2, which is an extraordinarily small number; to state this in a way that is independent of the arbitrary units adopted by experimentalists we note that the gravitational attraction between two electrons at rest is less than the electrostatic repulsion by a factor of about 4×10^{42}. In an attempt to understand the size of κ, among other things, Brans and Dicke (1961) have developed a theory in which κ is considered to be related to a new scalar field which is determined by the distribution of mass-energy in the universe. This is in accord with the ideas of Mach, who felt that the bulk material of the universe should somehow determine the inertial and hence the gravitational properties of individual bodies.

Let us motivate the introduction of the scalar field in the Brans-Dicke theory by noting an interesting numerical relation; as we shall discuss in Sec. 12.1, the characteristic "size" of the universe is about $R = 10^{10}$ light years, and its average density is very roughly 10^{-31} g/cm^3, with an uncertainty of several orders of magnitude. This leads to the very rough numerical relation

$$\frac{\kappa}{c^2}\frac{M}{R} \sim 1 \tag{11.124}$$

where M is the total mass of the universe (see Exercise 11.3). This may also be expressed as

$$\frac{1}{\kappa} \sim \frac{M}{c^2 R} \tag{11.124'}$$

The form of this relation suggests that $1/\kappa$ may be equated with a scalar field φ which is itself determined by a Poissonlike equation with the bulk matter density of the universe as source. That is,

$$\nabla^2\varphi \sim \frac{\rho}{c^2} \tag{11.125}$$

This clearly has $\varphi \sim M/Rc^2$ as a solution.

It is very easy to make these considerations precise and covariant. We replace $\nabla^2\varphi$ by $\varphi_{|\mu\|\nu}g^{\mu\nu}$ and the density ρ by the scalar $T^\alpha{}_\alpha$ and write

$$\varphi^{\|\mu}{}_{\|\mu} = \varphi_{\|\mu\|\nu}g^{\mu\nu} = \frac{4\pi\lambda T^\alpha{}_\alpha}{c^2} \qquad \varphi_{|\mu} = \varphi_{\|\mu} \tag{11.126}$$

The constant λ is a new dimensionless coupling constant. We anticipate that this new constant, which replaces κ as the fundamental constant of gravitational theory, will be of order unity, in accord with (11.125). That is, we adopt the attitude that a dimensionless constant of order unity is more "natural" or easily accepted than the dimensional constant κ, which is small in the sense noted above. Of course it is the task of experimental or observational physics, assuming the correctness of the theory, to determine the value of λ using the further development and predictions of the theory.

Equation (11.126) is the first fundamental equation of the Brans-Dicke theory. To preserve the structure of conventional relativity theory as much as possible we shall retain the general relativistic interpretation of the metric as determining the trajectories of test bodies in a curved Riemannian manifold; the φ field is to have no direct influence on such motion. This implies, for example, that $T^{\mu\nu}$ will be divergenceless, as discussed in Sec. 11.1. It thus remains only to obtain field equations for the metric tensor. We approach this problem by suitably generalizing the Lagrangian formulation of general relativity contained in (11.111). We rewrite it, with κ in evidence, as

$$\delta\left(\frac{J}{\kappa}\right) = \delta\int\left(\frac{1}{\kappa}R + \frac{8\pi}{c^2}L\right)\sqrt{-g}\,d^4x = 0 \tag{11.127}$$

The substitution for the constant $1/\kappa$ of the field φ is the obvious way to generalize the first term; the second term needs no modification, and L will be defined as before in (11.110). To include the dynamics of the scalar field we must add a suitable Lagrangian for φ. In special relativity it is easy to show that an appropriate Lagrangian is proportional to $\varphi_{|\alpha}\varphi_{|\beta}g^{\alpha\beta}$ (Exercise 11.4). In order to add such a term to the above Lagrangian without introducing a new dimensional constant we are led to

$$\delta J = \delta\int\left[\varphi R + \frac{8\pi}{c^2}L + \omega\frac{\varphi_{|\alpha}\varphi_{|\beta}}{\varphi}g^{\alpha\beta}\right]\sqrt{-g}\,d^4x = 0 \tag{11.128}$$

where ω is a dimensionless constant. We shall presently relate it to λ.

This generalization of (11.127) is clearly dimensionally consistent and evidently as simple as possible.

The variational problem (11.128) with φ considered as an independent field leads immediately to the Euler-Lagrange equations; using formula (3.12) for the divergence, we obtain

$$\frac{-2\omega}{\varphi}\,\varphi^{\|\alpha}{}_{\|\alpha} + \frac{\omega}{\varphi^2}\,\varphi_{\|\alpha}\varphi^{\|\alpha} + R = 0 \tag{11.129}$$

We next must vary $g^{\mu\nu}$ to obtain the remaining field equations. From (11.128), using (11.32) and (11.110), we obtain easily

$$\begin{aligned} \delta J = \int \Big[\varphi G_{\mu\nu} &+ \frac{8\pi}{c^2}\,T_{\mu\nu} + \frac{\omega}{\varphi}\,\varphi_{|\mu}\varphi_{|\nu} \\ &- \frac{\omega}{2\varphi}\,g_{\mu\nu}\varphi^{\|\alpha}\varphi_{|\alpha}\Big]\sqrt{-g}\,\delta g^{\mu\nu}\,d^4x \\ &+ \int \varphi g^{\mu\nu}\left[\delta\begin{Bmatrix}\alpha\\ \mu\ \alpha\end{Bmatrix}_{\|\nu} - \delta\begin{Bmatrix}\alpha\\ \mu\ \nu\end{Bmatrix}_{\|\alpha}\right]\sqrt{-g}\,d^4x \end{aligned} \tag{11.130}$$

Only the last two terms, which we shall label $\delta\mathfrak{M}$ and $\delta\mathfrak{N}$ must be further simplified to obtain a field equation linking $G_{\mu\nu}$ to $T_{\mu\nu}$ and φ, analogous to the Einstein equations. To simplify $\delta\mathfrak{M}$ we use (11.34) and integrate by parts, again using the divergence formula (3.12),

$$\begin{aligned} \delta\mathfrak{M} &= \int \varphi\left(g^{\mu\nu}\,\delta\begin{Bmatrix}\alpha\\ \mu\ \alpha\end{Bmatrix}\right)_{\|\nu}\sqrt{-g}\,d^4x \\ &= \int \varphi\left(\sqrt{-g}\,g^{\mu\nu}\delta\begin{Bmatrix}\alpha\\ \mu\ \alpha\end{Bmatrix}\right)_{|\nu} d^4x \\ &= -\int \varphi_{|\nu}g^{\mu\nu}\delta\begin{Bmatrix}\alpha\\ \mu\ \alpha\end{Bmatrix}\sqrt{-g}\,d^4x \end{aligned} \tag{11.131}$$

Finally we use (11.14), (11.30), and the fact that the variation operation and ordinary differentiation commute to obtain

$$\begin{aligned} \delta\mathfrak{M} &= -\int\varphi_{|\nu}g^{\mu\nu}\sqrt{-g}\,\delta\,(\log\sqrt{-g})_{|\mu}\,d^4x \\ &= \int(\varphi_{|\nu}g^{\mu\nu}\sqrt{-g})_{|\mu}\,\delta(\log\sqrt{-g})\,d^4x \\ &= -\tfrac{1}{2}\int\varphi^{\|\nu}{}_{\|\mu}g_{\mu\nu}\,\delta g^{\mu\nu}\sqrt{-g}\,d^4x \end{aligned} \tag{11.132}$$

which we see by reference to (11.130) is in convenient form.

The last term $\delta\mathfrak{N}$ involves a bit more algebra than the above but is straightforward. Proceeding as with $\delta\mathfrak{M}$, we simplify it to

$$\delta\mathfrak{N} = \int \varphi_{|\alpha}g^{\mu\nu}\delta\begin{Bmatrix}\alpha\\ \mu\ .\nu\end{Bmatrix}\sqrt{-g}\,d^4x \tag{11.133}$$

We observe that, analogous to the product rule for differentiation,

$$g^{\mu\nu}\delta\begin{Bmatrix}\alpha\\ \mu\ \nu\end{Bmatrix} = g^{\mu\nu}\,\delta g^{\alpha\sigma}[\mu\nu,\sigma] + g^{\mu\nu}g^{\alpha\sigma}\delta[\mu\nu,\sigma] \tag{11.134}$$

If we substitute this into (11.133) and integrate the last term by parts, the result is

$$\begin{aligned} \delta\mathfrak{N} = \int \Big[\varphi_{|\alpha}g^{\mu\nu}[\mu\nu,\sigma]\,\delta g^{\alpha\sigma} &- \varphi^{|\sigma}{}_{|\nu}g^{\mu\nu}\,\delta g_{\mu\sigma} - \varphi^{|\sigma}g^{\mu\nu}{}_{|\nu}\,\delta g_{\mu\sigma} \\ &- \varphi^{|\sigma}g^{\mu\nu}\begin{Bmatrix}\alpha\\ \alpha\ \nu\end{Bmatrix}\delta g_{\mu\sigma} + \tfrac{1}{2}\varphi^{\|\sigma}{}_{\|\sigma}g^{\mu\nu}\,\delta g_{\mu\nu} \\ &+ \tfrac{1}{2}\varphi^{|\sigma}g^{\mu\nu}{}_{|\sigma}\,\delta g_{\mu\nu}\Big]\sqrt{-g}\,d^4x \end{aligned} \tag{11.135}$$

To put this in a form analogous to (11.130) and (11.132) we use (11.28) and the following relations, which follow from the constancy of $g_{\mu\nu}g^{\nu\tau} = \delta_\mu{}^\tau$:

$$\begin{aligned} g^{\mu\nu}\,\delta g_{\mu\sigma} &= -g_{\mu\sigma}\,\delta g^{\mu\nu} \qquad \delta g_{\mu\nu} = -g_{\mu\alpha}g_{\gamma\beta}\,\delta g^{\alpha\beta} \\ g^{\omega\tau}{}_{|\mu}g_{\omega\sigma} &= -g^{\omega\tau}g_{\omega\sigma|\mu} \end{aligned} \tag{11.136}$$

This leads to the final covariant result for $\delta\mathfrak{N}$

$$\delta\mathfrak{N} = \int[-\tfrac{1}{2}\varphi^{\|\sigma}{}_{\|\sigma}g_{\mu\nu} + \varphi_{\|\mu\|\nu}]\,\delta g^{\mu\nu}\sqrt{-g}\,d^4x \tag{11.137}$$

We can now combine (11.130), (11.132), and (11.137) to obtain the following replacement for the Einstein field equations

$$\begin{aligned} R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = G_{\mu\nu} = -\frac{8\pi}{\varphi c^2}\,T_{\mu\nu} &- \frac{\omega}{\varphi^2}\,(\varphi_{|\mu}\varphi_{|\nu} - \tfrac{1}{2}g_{\mu\nu}\varphi^{|\alpha}\varphi_{|\alpha}) \\ &+ \frac{1}{\varphi}\,(\varphi_{\|\mu\|\nu} - g_{\mu\nu}\varphi^{\|\sigma}{}_{\|\sigma}) \end{aligned} \tag{11.138}$$

This equation also allows us to obtain the scalar R in terms of the scalars T and φ

$$R = \frac{8\pi}{c^2\varphi} T - \frac{\omega}{\varphi^2} \varphi_{|\mu}\varphi^{|\mu} - \frac{3}{\varphi} \varphi^{\|\alpha}{}_{\|\alpha} \tag{11.139}$$

Finally, substituting this into (11.129), we obtain a more useful form for the scalar field equation, which in fact is precisely that anticipated in (11.126)

$$\varphi^{\|\alpha}{}_{\|\alpha} = \frac{8\pi T}{c^2(3 + 2\omega)} \tag{11.140}$$

where we identify $\lambda = 2/(3 + 2\omega)$.

Equations (11.138) and (11.140) constitute a complete basis for the Brans-Dicke theory. One may investigate all the usual problems of general relativity theory, such as those discussed in Chap. 6, and make predictions which will differ from those of conventional general relativity theory, thereby providing a means of testing the theory and measuring the constant ω. The reader is referred to the problems at the end of this chapter and to the bibliography for details of the specific problems. Although we shall not discuss the specific predictions of the Brans-Dicke theory in detail, we note that in the limit of large ω or small λ (11.140) gives

$$\varphi^{\|\mu}{}_{\|\mu} = 0 \qquad \varphi = \text{const} = \langle\varphi\rangle \tag{11.141}$$

and so (11.138) becomes the usual Einstein equation with $\kappa = 1/\langle\varphi\rangle$. That is, the theory goes over in this limit to conventional general relativity (see also Exercise 11.5).

In order to obtain a lower estimate for ω we mention one specific prediction of the Brans-Dicke theory. In Chap. 6 we discussed the perihelion precession of Mercury and Dicke's suggestion that roughly 8 per cent is due to the quadrupole moment of the sun, leaving only 40″ to be explained in terms of relativistic effects. The Brans-Dicke theory predicts for this shift $(3\omega + 4)/(3\omega + 6)$ times the Einstein value. Thus the Brans-Dicke theory is in agreement with this result if we set $(3\omega + 4)/(3\omega + 6) = 0.92$, which implies $\omega \cong 6.2$, or $\lambda \cong 0.13$. Of course the quadrupole effect remains an unsettled question, as noted in Chap. 6. At present the other observational tests of relativity discussed in Chap. 6 are not capable of distinguishing between general relativity and the scalar tensor, but improvements in accuracy should provide definitive tests in the near future. Such tests will give a specific value for ω, or if they are consistent with general relativity, they will give a lower bound on ω. It is important to note that the existence of the scalar field cannot be disproved since general relativity is the limit case for $\omega = \infty$: it cannot be decided by observation whether ω is very large or infinite (Prob. 11.7).

Exercises

11.1 Discuss further the criteria for a test body noted in Sec. 11.1. In particular study how large a test body may be before gravitational tidal forces become appreciable and the body ceases to act as if it had zero size.

11.2 Verify (11.119).

11.3 The argument used to motivate the introduction of the scalar φ in the Brans-Dicke theory depends on the relation (11.124). However, a similar relation can be obtained as a consequence of conventional relativity theory, as we shall see in Chap. 13. Does this weaken the motivation for the scalar tensor theory?

11.4 Show that in special relativity a Lagrangian proportional to $\varphi_{|\alpha}\varphi_{|\beta}g^{\alpha\beta}$ leads to the wave equation $\Box^2\varphi = 0$ and is thus appropriate to a scalar field.

11.5 For large ω show that the Brans-Dicke equations (11.138) and (11.140) become

$$\varphi = \langle\varphi\rangle + O\left(\frac{1}{\omega}\right) = \frac{1}{\kappa} + O\left(\frac{1}{\omega}\right)$$

$$G_{\mu\nu} = -\frac{8\pi\kappa}{c^2} T_{\mu\nu} + O\left(\frac{1}{\omega}\right)$$

which explicitly illustrates the large ω limit.

11.6 Write the Brans-Dicke equation (11.138) as

$$G_{\mu\nu} = -\frac{8\pi}{c^2\varphi}(T_{\mu\nu} + B_{\mu\nu})$$

where $B_{\mu\nu}$ is interpreted as the energy-momentum tensor of the scalar field. Show that $B_{\mu\nu}$ is divergenceless from its definition.

Problems

11.1 The derivation of the geodesic equation of motion presented in the text concerned a structureless body, a small dust globule. Consider a small spinning body and obtain an equation of motion (see Papapetrou, 1951).

11.2 Consider an extended body and discuss corrections to the geodesic equation of motion. How might one study the motion of two bodies of comparable mass interacting gravitationally? What would be the effect of gravitational radiation on the motion?

11.3 The energy-momentum of a gravitating system was obtained as a surface integral in (11.102). A similar procedure can be used to obtain the angular momentum of a gravitating system (Cohen, 1968). Use this approach to identify the parameter a in the Kerr metric, as discussed in Sec. 7.7.

11.4 Obtain the energy momentum tensor of the gravitational field using canonical field theory (see Bjorken and Drell, 1965, for a discussion of the canonical theory in flat space).

11.5 Obtain the exterior field of a spherically symmetric body in Brans-Dicke theory, analogous to the Schwarzschild solution.

11.6 Obtain expressions for the planetary perihelion shift, the solar deflection of light, and time delay of radar pulses for the Brans-Dicke theory, in analogy with the results of Chap. 6 for the Schwarzschild solution. What of the gravitational red shift prediction? (See Weinberg, 1972.)

11.7 Compare the Brans-Dicke expressions from the above with the general relativistic expressions and with the observational values discussed in Chap. 6. What range of values may ω take for each test if the scalar tensor theory is to be consistent with observation?

Bibliography

Bergmann, P. G. (1942): "An Introduction to the Theory of Relativity," New York.

Brans, C., and R. H. Dicke (1961): Mach's Principle and a Relativistic Theory of Gravity, *Phys. Rev.*, **124**:925.

Bjorken, J. D., and S. D. Drell (1965): Relativistic Quantum Fields, New York.

Cohen, J. M. (1968): Angular Momentum and the Kerr Metric, *J. Math. Phys.*, **9**:905.

Einstein, A., and J. Grommer (1927): Allgemeine Relativitätstheorie und Bewegungsgesetz, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 2–13.

Einstein, A., and L. Infeld (1940): The Gravitational Equations and the Problem of Motion, II, *Ann. Math.*, **41**(2):455–464.

Einstein, A., L. Infeld, and B. Hoffmann (1938): The Gravitational Equations and the Problem of Motion, *Ann. Math.*, **39**(2):pp. 65–100.

Fock, V. (1959): "The Theory of Space, Time and Gravitation," New York.

Infeld, L., and J. Plebanski (1960): "Motion and Relativity," London.

Levi-Civita, T. (1964): "The N-Body Problem in General Relativity," Dordrecht, chap. 1.

Palatini, A. (1919): Deduzioni invariantive delle equazioni gravitazionali dal principio di Hamilton, *Rend. Circ. Mat. Palermo*, **43**:203–212.

Papapetrou, A. (1951): Spinning Test Particles in General Relativity I, *Proc. Roy. Soc.*, **209**:248.

Trautman, A. (1962): Conservation Laws in General Relativity, in L. Witten (ed.) "Gravitation: An Introduction to Current Research," New York, pp. 169–198.

Weinberg, S. (1972): "Gravitation and Cosmology," New York, chap. 7.

See also the standard books on general relativity theory.

CHAPTER

12

Descriptive Cosmic Astronomy

In the various phenomena which physicists investigate a small number of fundamentally different types of interactions occur; we observe the effects of the "strong" and "weak" nuclear interactions, the electromagnetic interactions, and the gravitational interactions. Of these the gravitational interactions are by far the weakest. Nevertheless, because the nuclear interactions are of short range and large aggregates of matter tend to be electrically neutral, it appears that the large-scale phenomena of the universe are most strongly affected by the gravitational interaction. If gravity is indeed the dominating influence, then as a theory of gravity, general relativity should be able to give a description of the universe in the large. The cosmological problem within the framework of general relativity consists in finding a model of the universe as a whole which is a solution of Einstein's equations. Such a model is intended to describe *only* the large-scale state of the universe; for example, the energy-producing processes which take place inside stars are of nuclear origin and clearly cannot be described by relativity theory. Moreover, the process of galactic evolution, although largely determined by the gravitational interaction, is on a much smaller scale than we wish to consider and must be separated from the cosmological problem. An acceptable cosmological model need therefore fit only a limited number of observational facts of a global nature. We shall first discuss the observational facts that a cosmological model should describe, and in particular the presently determined values of some important cosmological parameters. In the rest of the chapter we shall discuss some general features of the cosmological problem, preparatory to discussing specific cosmological models in the next chapter.

12.1 Observational Background

In astronomical studies one characterizes a star or galaxy by position coordinates. These provide a marker system in the non-Euclidean space which we use for theoretical study. In particular, one attaches to a star or galaxy a radial distance marker referred to as its *astronomical distance*. For the nearest stars the astronomical distance may be determined by triangulation, using local Euclidean geometry; the base line is the size of the earth's orbit, and the angle is the star's parallax. This method is the oldest and simplest but is limited to a few thousand stars within about 100 light years, since more distant stars have too small a parallax to measure accurately. In particular it is certainly limited to stars in our own galaxy and cannot directly yield information of cosmological interest (see Prob. 12.1, however).

For more distant stars we must proceed differently. To do this we make use of the inverse-square law of decreasing light intensity. Suppose we know that a distant star (or galaxy) has the same total rate of emission, or *absolute luminosity*, as another closer star whose distance L_0 is known; calling E and E_0 the apparent luminosities of the two stars, we can define an astronomical distance L for the more distant star by the relation

$$\frac{L^2}{L_0^2} = \frac{E_0}{E} \tag{12.1}$$

It is known that the surface temperature (or equivalently the spectral type) and the absolute luminosity of the so-called *main-sequence* stars are correlated in a simple way. (This is usually shown on a plot of absolute luminosity versus surface temperature known as a *Hertzsprung-Russell diagram.*) This correlation allows one to use relation (12.1) in practice by comparing two stars of the same spectral type.

For larger distances certain variable stars may be used as markers; such stars possess an absolute luminosity that varies periodically in time, with a period that is directly related to their absolute luminosity. The most important of these are the classical Cepheid variables (Leavitt, 1912). Such stars can be seen in other galaxies. Indeed Cepheids led to the first confirmation of the extragalactic nature of spiral nebulae such as M31 in Andromeda, a galaxy about 2×10^6 light years away. The use of Cepheids as distance markers has been historically fraught with difficulty and confusion; e.g., there are two distinct types of Cepheids. It is hoped that the calibration problem has now been cleared up. This difficulty in the use of Cepheids is one of the reasons for the large changes in our estimates of the scale of the universe during the twentieth century.

For more distant galaxies, those outside the local cluster of galaxies in which we reside, we need brighter stars than the Cepheids. The distance markers used are the supergiant blue stars, or novae which flare up briefly to approximately known absolute magnitudes, or entire globular clusters of stars.

Finally, to obtain distance markers of cosmological significance one must use entire galaxies since no individual stars are sufficiently bright. For this purpose certain types of galaxies whose absolute luminosities appear to be well defined are used. One representative method is to note that galaxies tend to occur in clusters of hundreds or thousands; it is then reasonable to suppose that the brightest galaxies (generally the elliptical, or E, galaxies) in various clusters all have approximately the same absolute luminosity, which provides a reference luminosity. A more detailed discussion is given by Hubble (1957) and by Sandage (1961), and a briefer summary by Weinberg (1972).

In summary, the procedure used to define the astronomical distance L is to extend all over the world a radial *marker system L in which the law of energy decrease with the inverse square of the distance is strictly true* and which coincides with the radial coordinate used in our local Euclidean geometry at small distances. The distance marker L thus defined is sometimes called an "energy distance."

At the beginning of Chap. 2, we mentioned briefly the concept of a Riemannian manifold which is pieced together by overlapping coordinate neighborhoods with known transformation laws for transition from one mapping of the manifold into another. The less mathematically inclined reader might have considered our definition as an axiomatical-logical refinement of no great practical value in experimental science. It may therefore be illuminating to point out at this stage how close observational and measuring procedures are to this concept of overlapping coordinate sets. In the above discussion we mention first the neighborhood of the earth, where distances can be measured by parallaxes and angular readings. In this neighborhood we obtain the main-sequence correlation referred to above and extend the distance scale to where we are able to gauge the Cepheid variables and establish their luminosity-periodicity law. Then these variable stars can be used as markers far beyond the first coordinate map. In the new piece of the cosmos opened up for measurement, we discover the spectral-type laws which enable us by a further extension to formulate the Hubble distance–red shift relation discussed below, which opens up new coordinate possibilities. The mathematical concept of a pieced-together manifold is precisely a formulation of these classical astronomical practices.

While determining the apparent luminosity of stars, astronomers also determined and classified their line spectra and discovered that a shift of the known spectral lines toward the red occurred for stars in galaxies which were farther and farther away. For each distant galaxy, the relative shift in wavelength $\Delta\lambda/\lambda$ appeared to be proportional to the astronomical distance of the galaxy as defined above. We shall see later that this striking phenomenon can be successfully interpreted as a recession of distant galaxies whose speed increases linearly with their distance from us. One can then visualize the universe as *expanding*. Most of the original astronomical observations of this effect are due to Hubble (1936), and the results can be summarized by the simple relation

$$\frac{\Delta\lambda}{\lambda} \cong \frac{L}{c} H \tag{12.2}$$

where the approximate equality sign reflects experimental uncertainty. This relation is known as Hubble's law, and the important number H is called Hubble's constant. Its first evaluation in 1936 gave a value $H^{-1} = 0.56 \times 10^{17}$ s, which was used in the literature until 1958. This number, which may be interpreted as a fundamental distance or time scale of the universe, has undergone considerable change since then. Sandage pointed out in 1958 that Hubble had apparently confused interstellar hydrogen clouds with individual blue stars in his original estimate of H. Indeed, if this and all known sources of error and uncertainty are combined, the value of H should be revised downward by nearly a factor of 10 and assigned a rather large uncertainty. The value of H in present use is (Sandage, 1972)

$$H^{-1} = (5.6 \pm 0.6) \times 10^{17} \text{ s} = (1.8 \pm 0.2) \times 10^{10} \text{ years}$$

$$\frac{\Delta\lambda}{\lambda} = \frac{L'}{c} H \tag{12.3}$$

with a given fixed value of H is strictly true. This system of markers, called *Hubble distances*, can be extended to very remote galaxies for which an "energy distance" is very difficult to estimate accurately because of very low apparent luminosity and because of selective light absorption in intergalactic space. A Hubble distance, on the other hand, can be obtained very accurately because the shift in wavelength can be measured with high precision as soon as one spectral line is identified.

Very interesting objects, called *quasi-stellar objects* or *quasars*, many of which have extraordinary red shifts, have lately been observed. The greatest red shift reported so far is $\Delta\lambda/\lambda = 3.5$, which is so large that time-dilation effects are important and (12.3) should not be used to

define its distance (see Chap. 4). It is clear that if this red shift is actually caused by the cosmological expansion of the universe, the quasar must be far away indeed, about 95 per cent of the way to the "edge" of the universe, i.e., the distance at which the galaxies recede from us at the velocity of light and are unobservable. Such an object is obviously of great cosmological interest. Unfortunately the physical nature of the quasars is not yet understood; since they vary in luminosity in typical times of about a month they cannot be much larger than a light month in size, which is very much smaller than the 10,000–light year size of a typical galaxy. On the other hand their absolute luminosity appears to range from 1 to 100 times that of a typical galaxy. The question of how such a small object can radiate so much energy has presented theoreticians with very interesting and challenging problems which are not yet solved. The role of quasars in cosmology at the present time is uncertain, but their study may lead to drastic changes in our concepts of the evolution of galaxies, the early universe, and cosmology (see Morrison, 1973).

To apply Einstein's equations to the actual universe with a precise physical distribution of matter represented by an exact energy-momentum tensor $T^{\mu\nu}$ would obviously be a hopeless mathematical task. In order to achieve a tractable description of the matter distribution in the universe in the large, theoretical cosmologists always make the idealizing assumption that, on a sufficiently large scale, matter can be considered to be homogeneously distributed. Observationally, out to the largest distances reached by present-day telescopes, this uniformity appears on the scale of the clusters of galaxies, but not on the scale of individual galaxies whose clustering tendencies are quite evident (Oort, 1958). One may then consider an idealized universe of uniform continuous matter distribution represented by a constant density of matter energy ρ_0 which serves as the $T^0{}_0$ component of the energy-momentum tensor.

Estimates of the average density of the universe are very difficult. To obtain the density due to the ponderable matter in galaxies one must count galaxies out to some Hubble distance, divide by the volume in which they are contained, and multiply by their average mass. The determination of the volume requires a knowledge of Hubble's constant, which is not known with great accuracy. To obtain the masses of galaxies one may, for example, analyze the relative velocities and separations of a pair of galaxies, a process which is subject to considerable uncertainty. Using such methods it is estimated (Oort, 1958) that $\rho_0 = 2 \times 10^{-31}$ g/cm³. This number could easily be in error by a factor of 3. In addition to the ponderable matter in galaxies ρ_0 should contain the density of intergalactic dust and gas (Beer, 1960), dim intergalactic stars, black holes, particles such as cosmic rays, neutrinos, the quanta of gravitational radiation (called *gravitons*), photons, and possibly others. It is difficult to estimate the density of such material or even to get a reasonable upper limit. For example, the only upper limit on the neutrino density that would result from inverse β-decay reactions in stars is several orders of magnitude greater than the above estimate for the galactic ponderable mass (Pontecorvo and Smorodinsky, 1962). As a result it is impossible at present to place a safe upper limit on ρ_0; it could easily be as large as 10^{-28} g/cm^3. As we shall discuss later in Sec. 13.3, there is observational evidence which, when combined with theory, suggests that the mass density should be about 0.5×10^{-29} g/cm^3 or greater.

The field equations for a space filled homogeneously with matter lead to an evolution of the universe in time, and this evolution can be compared with astronomical observations. Some models correspond to a static universe (all of which are unstable, however) and some to a dynamic universe with limited life span. In the latter case, there is an origin of time, and the time elapsed to attain the present state of the universe should be compatible with experimental estimates of "ages," or periods of evolution, of different parts of the universe. The age of the universe should, for instance, be larger than the age of the earth and the solar system, which can be determined from the ratios of abundances of radioactive substance to decay products in various rocks and meteorites. This method leads to an estimate of the earth's age as $(4.55 \pm 0.07) \times 10^9$ years (Patterson, 1960). The age of the universe should also be greater than or equal to the age of the oldest stars. When plotting the usual color-absolute magnitude diagram for stars in our galaxy, one finds some stars that are evidently 10^{10} years old (Wilson, 1950); for other samples of stars, this same age and ages up to 3.2×10^{10} years have been found (Oke, 1959).

The fact that relativity predicts an evolutionary universe was actually recognized by Einstein but rejected as physically unreasonable. He then introduced a "cosmological constant" to obtain a static universe (Sec. 13.2). Only later was it discovered that the universe does indeed seem to be expanding and evolving, thus making the ad hoc introduction of the cosmological constant unnecessary. Further evidence for the evolutionary nature of the universe was obtained in 1965 by Penzias and Wilson (1965). They discovered that space appears to be filled with electromagnetic radiation with a blackbody spectrum, implying that it is in thermal equilibrium, with a temperature at present of about 3°K. Dicke and collaborators (1965) suggested that this radiation is actually the cooled remnant of a primordial fireball that accompanied the explosive birth of the universe about 2×10^{10} years ago, a time equal

to the inverse of Hubble's constant. The consequences of such a "big bang" birth of the universe had been investigated by Gamow, Alpher, and others many years before (Gamow, 1946; Alpher, 1948). Further measurements of the radiation intensity confirmed the character of the blackbody spectrum, as summarized by Dicke (1970), thus lending strong support to the validity of an expanding-universe concept with a big-bang birth. Conversely it is evidence against the steady-state model, which we discuss in Sec. 13.5.

12.2 The Mathematical Problem in Outline

The mathematical task of solving the cosmological problem consists in determining a large-scale metric of the four-dimensional world and a corresponding large-scale mass-energy distribution satisfying Einstein's equations. The metric, in turn, gives physical predictions since the galaxies and light rays move along geodesics of the four-dimensional space which it defines. In the present case, one has to solve a problem in the large in contrast to the local Schwarzschild problem treated earlier; in the Schwarzschild case, a metric was determined from a priori knowledge of the energy-momentum tensor, which was zero except at one single point; in the present case the metric and the energy density have to be determined together over the four-dimensional world. When the cosmological problem is solved, it will give the large-scale average behavior of the physical world, and therefore give the boundary condition to be used at infinity for local models like the Schwarzschild solution.

To restrict the possible forms of a cosmological metric, we shall first impose the requirement of spatial isotropy, which we have seen appears to be physically justifiable on a large enough scale. Then different models can be developed corresponding to either static or time-dependent solutions and to different values of a characteristic parameter which will appear in the general form of the isotropic metric and is related to the curvature of the three-dimensional world. From each such model one can derive properties of the corresponding physical world. A comparison of these predictions with our observational knowledge should provide a test of the validity of the model. Unfortunately, the general imprecision of the experimental astronomical data relevant to cosmological problems tends to make one consider any attempt at comparison with theoretical models easy or difficult, according to one's individual scientific integrity.

We have mentioned, in Chap. 10, that the most general form of Einstein's equations under certain reasonable mathematical requirements is

$$R^{\nu}{}_{\mu} - \tfrac{1}{2}Rg^{\nu}{}_{\mu} + \Lambda g^{\nu}{}_{\mu} = -\frac{8\pi\kappa}{c^2}T^{\nu}{}_{\mu} \tag{12.4}$$

where Λ must be a constant, the so-called cosmological constant. We solved these equations earlier (Chap. 6), neglecting the cosmological term when studying the Schwarzschild solution. This solution was seen to be in satisfactory agreement with the available observational tests within the solar system. In Chaps. 10 and 11 we also made investigations of these equations, assuming Λ to be zero. We shall now discuss the complete equations.

We notice at once that the general form of Einstein's equations (12.4) with $\Lambda \neq 0$ does not admit a flat-space solution for an empty universe. This is evident since an empty universe by definition is characterized by $T^\nu{}_\mu = 0$ and a flat space is characterized by $R^\nu{}_\mu = 0$; therefore Eqs. (12.4) cannot be satisfied unless $\Lambda = 0$. That is, an empty space cannot be described mathematically using Euclidean geometry unless $\Lambda = 0$.

In our previous treatment of problems in gravitational theory, however, we considered masses isolated in a finite part of the universe and tried to solve the field equations with the boundary condition of Euclidean geometry at infinity. This was based on the assumption that the influence of the finite-mass distribution is negligible at infinity and that in empty space Euclidean geometry prevails. In other words, we assumed then that an empty space is flat. Our theoretical predictions for phenomena within the solar system were in agreement with experiment. Therefore, on the scale of the solar system, the presence of a Λ-term in the general form of Einstein's equations must have a negligible influence; thus, if nonzero, Λ must at least be very small.

This can also be expected from dimensionality considerations. In any system of coordinates x^μ which have the dimensions of length, the $g_{\mu\nu}$ are dimensionless from the definition (1.2) of a line element. The elements of the contracted Riemann tensor then have the dimension of inverse length squared since they are obtained from the $g_{\mu\nu}$'s by double differentiation. To preserve homogeneity in (12.4) it is therefore necessary for Λ to have the dimension of the square of an inverse length. Indeed, we shall see presently that the integration of (12.4) for an empty space leads to a finite universe with a "characteristic size" of $1/\sqrt{\Lambda}$. On dimensional grounds one might expect from this that Λ should in general be related to the inverse square of the size of the universe and should therefore play a negligible role in local problems.

Let us digress for a moment to show that this statement can be illustrated in a simple way by considering the Schwarzschild problem with the inclusion of the cosmological term $\Lambda g_{\mu\nu}$. Setting $T_\mu{}^\nu = 0$ in (12.4), we find by contraction that $R = 4\Lambda$, and hence that

$$R_{\mu\nu} = \Lambda g_{\mu\nu} \tag{12.5}$$

Let us briefly indicate how to solve this equation in the case of a static radially symmetric metric as given in (6.9) and (6.26). We observe that the expression for the contracted Riemann tensor $R_{\mu\nu}$ is the same as in our calculations of Chap. 6. We conclude from (6.31) and the equation $R_{00} = \Lambda g_{00}$ that

$$\nu'' + \tfrac{1}{2}\nu'^2 - \tfrac{1}{2}\nu'\lambda' + \frac{2\nu'}{r} = -2\Lambda e^\lambda \tag{12.6}$$

and from (6.35) and $R_{11} = \Lambda g_{11}$,

$$\nu'' + \tfrac{1}{2}\nu'^2 - \tfrac{1}{2}\nu'\lambda' - \frac{2\lambda'}{r} = -2\Lambda e^\lambda \tag{12.7}$$

Thus subtraction of (12.7) from (12.6) yields, as in the original Schwarzschild problem,

$$\nu' + \lambda' = 0 \qquad \nu + \lambda = \text{const} = \log k \tag{12.8}$$

We have, however, no possibility of determining the value of this constant since we cannot ask for a Lorentzian character of the solution at infinity. We write, therefore,

$$\lambda = \log k - \nu \tag{12.9}$$

and insert into (12.6) to obtain

$$(e^\nu)'' + \frac{2}{r}(e^\nu)' = \frac{1}{r}(re^\nu)'' = -2\Lambda k \tag{12.10}$$

This yields the integral

$$re^\nu = \alpha + \beta r - \Lambda\frac{kr^3}{3} \tag{12.11}$$

with α, β as constants of integration.

Next we use (6.44) and the field equation $R_{22} = \Lambda g_{22}$ to obtain

$$\frac{\partial^2}{\partial\theta^2}(\log|\sin\theta|) + (e^{-\lambda}r)' - 2e^{-\lambda} + \cot^2\theta + e^{-\lambda}r\left(\frac{\lambda' + \nu'}{2} + \frac{2}{r}\right) = -\Lambda r^2 \tag{12.12}$$

Rearranging and using (12.8), we reduce this to

$$(e^{-\lambda}r)' = 1 - \Lambda r^2 \tag{12.13}$$

which, by virtue of (12.9), is the same as

$$(re^{\nu})' = k(1 - \Lambda r^2) \tag{12.14}$$

Comparing (12.11) and (12.14), we conclude that

$$k = \beta \tag{12.15}$$

Thus, from (12.9) and (12.11), it follows that

$$e^{\nu} = k\left(1 + \frac{\alpha}{kr} - \Lambda\frac{r^2}{3}\right) \qquad e^{-\lambda} = \left(1 + \frac{\alpha}{kr} - \Lambda\frac{r^2}{3}\right) \tag{12.16}$$

In view of the form of the line-element (6.9) we can always choose our unit of time such that $k = 1$, and therefore

$$e^{\nu} = 1 + \frac{\alpha}{r} - \Lambda\frac{r^2}{3} = e^{-\lambda} \tag{12.17}$$

Thus our resulting solution is identical with (6.53) except that the term $(1 - 2m/r)$ has to be replaced by $(1 - 2m/r - \frac{1}{3}\Lambda r^2)$.

We remark, first, that in this solution we do not obtain the ordinary Lorentzian metric if we suppose it to be regular everywhere, i.e., if we assume $m = 0$. In order to interpret the geometry of the line element which belongs to a radially symmetric static solution that is regular at the center (that is, with $m = 0$), let us consider the spatial part of the line element

$$dl^2 = \frac{dr^2}{1 - \frac{1}{3}\Lambda r^2} + r^2\,d\theta^2 + r^2 \sin^2\theta\,d\varphi^2 \tag{12.18}$$

As we shall see in the next section [formula (12.62)], this line element corresponds to the metric on the three-dimensional surface of a four-dimensional hypersphere with radius $R = \sqrt{3/\Lambda}$. This shows clearly the relation between Λ and the size R of the finite universe described by (12.18). Obviously, for r much smaller than the size of the universe, the presence of the extra term $\frac{1}{3}\Lambda r^2$ in the Schwarzschild solution may be neglected. Thus observation within our planetary system would not enable us to determine the presence or absence of the Λ-term in the field equations.

We shall later discuss various models of isotropic worlds which are empty or filled homogeneously with matter. It should be pointed out here that the present Schwarzschild solution with $m = 0$ does not describe an empty isotropic model of this kind. While we can easily attain spatial isotropy by a proper change of the radial variable, the center of radial symmetry always plays a distinguished role in the time term of the line element. Furthermore, there occurs a radial force directed away from this center of symmetry, which we can easily demonstrate explicitly. For small values of m we may use the perturbation theory of Sec. 4.3 to interpret the dynamical consequences of the Schwarzschild metric in the case of a nonvanishing cosmological term Λ. We have, on the one hand,

$$g_{00} = 1 - \frac{2m}{r} - \tfrac{1}{3}\Lambda r^2 \tag{12.19}$$

and on the other, by (4.142),

$$g_{00} = 1 + \frac{2\varphi}{c^2} \tag{12.20}$$

where $\varphi(x^i)$ is the potential in classical mechanics which would induce motion of a test particle approximately along the geodesics of the metric considered. Thus we obtain the approximate correspondence

$$\varphi = -\frac{\kappa M}{r} - \tfrac{1}{6}\Lambda c^2 r^2 \tag{12.21}$$

Therefore, even in the absence of the mass M at the origin, a test particle would be subjected to a radial acceleration

$$a = \tfrac{1}{3}\Lambda c^2 r \tag{12.22}$$

This shows that the origin of the coordinate system is dynamically distinguished from all other points in space, and is not an arbitrary point in three-dimensional space.

One last remark about the solution (12.17) and the space line element (12.18) for $m = 0$ can be made. The solution could be realized by some sort of radially symmetric distribution of matter about the origin; the neighborhood of the origin where the solution holds must of course be free of matter. Thus one arrives at a notion of a spherical cavity inside some spherical distribution of mass as having a metric solution (12.17). It is thus easily seen that, for $\Lambda = 0$, one obtains a pseudo-Euclidean metric, while for $\Lambda \neq 0$, we obtain a non-Euclidean geometry (12.18). Thus in principle the acceleration (12.22) inside a mass shell could serve to measure the cosmological constant Λ.

From the foregoing example we see that the original requirement of flat space at infinity need not be retained when dealing with cosmological problems; we are still ignorant of the large-scale topology of physical three-dimensional space and of the four-dimensional space-time manifold. Thus it is clear that one cannot definitely decide whether or not the cosmological term $\Lambda g_{\mu\nu}$ in Einstein's equations should be set equal to zero. We shall present models for the two cases separately. However, in both cases we shall restrict from the beginning the functional form of the potentials $g_{\mu\nu}$ by certain postulates of symmetry which we shall discuss next.

12.3 The Robertson-Walker Metric

We mentioned earlier that the large-scale distribution of extragalactic nebular clusters in space is roughly isotropic around our own galaxy and that the number of clusters in a given volume seems to remain constant everywhere: the larger the scale considered, the better the approximation. To simplify the mathematical description, we therefore make the physical assumption that matter is distributed homogeneously in the world. Furthermore, we desire that the geometry of space be determined by the matter distribution. This general requirement is known as Mach's principle. It is the name given to it by Einstein (1918) when generalizing Mach's original hypothesis that the inertia of one body is due to the presence of all other bodies in the universe (Mach, 1883). Einstein's equations are one possible particularization of this principle: $T_{\mu\nu}$ characterizes the matter-energy distribution, and $G_{\mu\nu}$ the space geometry.

On the basis of Mach's principle, we shall require that the geometry of the three-dimensional space be homogeneous like the matter distribution. We shall thus start with a purely geometrical study of the form of a metric which describes a four-dimensional space containing a three-dimensional subspace of homogeneous geometry. The precise link with the physical assumption of homogeneity of the mass distribution will be elucidated afterwards through a specialization of the coordinate system used.

The following mathematical hypotheses are our starting point:

1. There exists a global time-coordinate which serves as the x^0 of a Gaussian coordinate system as defined in Chap. 2. [See Eq. (2.28).]
2. The three-dimensional spaces belonging to various constant values of this time-coordinate are locally isotropic.
3. Any two points in a three-space belonging to a given fixed time are equivalent.

Let us first remark that the use of a distinguished time-coordinate marks the abandonment of a completely covariant treatment of the cosmological problem. This is the price one has to pay to simplify the cosmological models and to describe physical reality in convenient mathematical terms. The fact that x^0 is a Gaussian time-coordinate means that, at some given moment, the matter of the universe is on the average at rest in its distinguished three-space. It therefore describes geodesic trajectories which are orthogonal to that three-space. As was shown in Chap. 2, these geodesics will remain parallel to the x^0 axis in all future three-spaces. That is, at any moment the matter of the universe will be at rest in this distinguished three-space. We may therefore refer to this coordinate system as the *co-moving system*. This point will be discussed in more detail in the next section.

The second condition is a local requirement, which is all that is necessary in most mathematical problems of general relativity theory. Let us consider an *arbitrary* but fixed point in space-time. As we discussed in Sec. 6.2, local isotropy implies that the space coordinates must appear in the line element ds^2 in the spherically symmetric combination

$$d\sigma^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \tag{12.23}$$

or equivalently, in r, θ, φ coordinates centered about the given point

$$d\sigma^2 = dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\varphi^2 \tag{12.24}$$

The complete line element in our Gaussian coordinates is therefore expressible as

$$ds^2 = (dx^0)^2 - e^{G(x^0,r)}\,d\sigma^2 \tag{12.25}$$

where the function G cannot depend on θ and φ because of the assumed isotropy of three-space around the given point. Writing ds^2 in the form (12.25), with an exponential function, enables us to keep the signature in evidence and guarantees that the determinant of $g_{\mu\nu}$ does not vanish.

Finally, the third postulate, the equivalence of all points in three-space at all times, requires that two observers at two different points observe a similar physics; the only thing which may differ between the two observers is the measuring scale they use. Therefore the ratio of physical or proper-distance elements at two points in space defined, respectively, by the coordinates (r_1,θ_1,φ_1) and (r_2,θ_2,φ_2) must remain fixed in time; that is, the ratio

$$\frac{\exp [G(x^0,r_1)]}{\exp [G(x^0,r_2)]} \tag{12.26}$$

must be independent of x^0. Therefore one must have

$$G(x^0,r_1) = G(x^0,r_2) + F(r_1,r_2) \tag{12.27}$$

If we now choose a fixed value of r_2, we can write

$$G(x^0,r_1) = g(x^0) + f(r_1) \tag{12.28}$$

We can therefore obtain the following form of the line element (12.25):

$$ds^2 = (dx^0)^2 - e^{g(x^0)+f(r)}\, d\sigma^2 \tag{12.29}$$

which is based entirely on symmetry arguments. The functions f and g which appear in the exponent are arbitrary (g is not to be confused with the determinant of $g_{\mu\nu}$).

To investigate the form of ds^2 still further, we consider now the Einstein equations (12.4). In order to calculate the Ricci tensor which enters into (12.4), we have to determine the Christoffel symbols of the metric (12.29). This is a straightforward calculation, but for the convenience of the reader, we shall indicate briefly the main steps in this computation. The geodesics of the metric (12.29) satisfy the variational condition

$$\delta\int[(\dot{x}^0)^2 - e^G(\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta\, \dot{\varphi}^2)]\, ds = 0 \tag{12.30}$$

if we use the coordinates x^0, r, θ, and φ. We obtain, therefore, the following Euler-Lagrange equations:

$$\begin{aligned}
&\ddot{x}^0 + \tfrac{1}{2}g'e^G(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\varphi}^2) = 0\\
&\ddot{r} + \tfrac{1}{2}f'\dot{r}^2 + g'\dot{x}^0\dot{r} - \left(\tfrac{1}{2}f' + \frac{1}{r}\right)(r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\varphi}^2) = 0\\
&\ddot{\theta} + 2\left(\tfrac{1}{2}f' + \frac{1}{r}\right)\dot{r}\dot{\theta} + g'\dot{x}^0\dot{\theta} - \sin\theta\cos\theta\,\dot{\varphi}^2 = 0\\
&\ddot{\varphi} + 2\left(\tfrac{1}{2}f' + \frac{1}{r}\right)\dot{r}\dot{\varphi} + g'\dot{x}^0\dot{\varphi} + 2\cot\theta\,\dot{\theta}\dot{\varphi} = 0
\end{aligned} \tag{12.31}$$

Here the dot denotes differentiation with respect to the parameter s, while $g'(x^0)$ and $f'(r)$ denote the derivatives of $g(x^0)$ and $f(r)$ with respect to their single arguments.

From (12.31) we can read off the list of nonvanishing Christoffel symbols in the metric (12.29). We find that

$$\begin{aligned}
&\begin{Bmatrix}0\\1\ 1\end{Bmatrix} = \tfrac{1}{2}g'e^G \qquad \begin{Bmatrix}0\\2\ 2\end{Bmatrix} = \tfrac{1}{2}g'e^Gr^2 \qquad \begin{Bmatrix}0\\3\ 3\end{Bmatrix} = \tfrac{1}{2}g'e^Gr^2\sin^2\theta\\
&\begin{Bmatrix}1\\0\ 1\end{Bmatrix} = \tfrac{1}{2}g' \qquad \begin{Bmatrix}1\\1\ 1\end{Bmatrix} = \tfrac{1}{2}f' \qquad \begin{Bmatrix}1\\2\ 2\end{Bmatrix} = -r^2\left(\tfrac{1}{2}f' + \frac{1}{r}\right)\\
&\qquad\qquad\qquad\qquad\qquad \begin{Bmatrix}1\\3\ 3\end{Bmatrix} = -r^2\left(\tfrac{1}{2}f' + \frac{1}{r}\right)\sin^2\theta\\
&\begin{Bmatrix}2\\0\ 2\end{Bmatrix} = \tfrac{1}{2}g' \qquad \begin{Bmatrix}2\\1\ 2\end{Bmatrix} = \left(\tfrac{1}{2}f' + \frac{1}{r}\right) \qquad \begin{Bmatrix}2\\3\ 3\end{Bmatrix} = -\sin\theta\cos\theta\\
&\begin{Bmatrix}3\\0\ 3\end{Bmatrix} = \tfrac{1}{2}g' \qquad \begin{Bmatrix}3\\1\ 3\end{Bmatrix} = \left(\tfrac{1}{2}f' + \frac{1}{r}\right) \qquad \begin{Bmatrix}3\\2\ 3\end{Bmatrix} = \cot\theta
\end{aligned} \tag{12.32}$$

From (12.29) we find for the determinant g of the metric tensor

$$\log\sqrt{-g} = \tfrac{3}{2}g(x^0) + \tfrac{3}{2}f(r) + 2\log r + \log|\sin\theta| \tag{12.33}$$

From the above it is then easy to see that

$$\begin{aligned}
&\begin{Bmatrix}\rho\\0\ 0\end{Bmatrix}_{|\rho} = 0 \qquad \begin{Bmatrix}\rho\\1\ 1\end{Bmatrix}_{|\rho} = \tfrac{1}{2}e^G(g'' + g'^2) + \tfrac{1}{2}f''\\
&\begin{Bmatrix}\rho\\2\ 2\end{Bmatrix}_{|\rho} = \left[\tfrac{1}{2}e^G(g'' + g'^2) - \left(\tfrac{1}{2}f'' + \frac{1}{r}f' + \frac{1}{r^2}\right)\right]r^2\\
&\begin{Bmatrix}\rho\\3\ 3\end{Bmatrix}_{|\rho} = \left[\tfrac{1}{2}e^G(g'' + g'^2) - \left(\tfrac{1}{2}f'' + \frac{1}{r}f'\right)\right]r^2\sin^2\theta - \cos^2\theta
\end{aligned} \tag{12.34}$$

which will be very useful shortly. We also need the following relations, which can be obtained from (12.32):

$$\begin{aligned}
&\begin{Bmatrix}\alpha\\0\ \rho\end{Bmatrix}\begin{Bmatrix}\rho\\0\ \alpha\end{Bmatrix} = \tfrac{3}{4}g'^2\\
&\begin{Bmatrix}\alpha\\1\ \rho\end{Bmatrix}\begin{Bmatrix}\rho\\1\ \alpha\end{Bmatrix} = \tfrac{1}{2}e^Gg'^2 + \tfrac{3}{4}f'^2 + \frac{2}{r}f' + \frac{2}{r^2}\\
&\begin{Bmatrix}\alpha\\2\ \rho\end{Bmatrix}\begin{Bmatrix}\rho\\2\ \alpha\end{Bmatrix} = \left[\tfrac{1}{2}e^Gg'^2 - \tfrac{1}{2}f'^2 - \frac{2}{r}f' - \frac{2}{r^2} + \frac{1}{r^2}\cot^2\theta\right]r^2\\
&\begin{Bmatrix}\alpha\\3\ \rho\end{Bmatrix}\begin{Bmatrix}\rho\\3\ \alpha\end{Bmatrix} = \left[\tfrac{1}{2}e^Gg'^2 - \tfrac{1}{2}f'^2 - \frac{2}{r}f' - \frac{2}{r^2} - \frac{2}{r^2}\cot^2\theta\right]r^2\sin^2\theta
\end{aligned} \tag{12.35}$$

This completes the chore of obtaining the various terms we shall need in expressing the field equations in terms of the functions f and g.

The most convenient way now to write the contracted Riemann tensor is to use the definition of $R_{\mu\nu}$ in (5.119) and the expression (3.11) for the contracted Christoffel symbol, $\begin{Bmatrix} \alpha \\ \beta\ \alpha \end{Bmatrix} = (\log\sqrt{-g})_{|\beta}$, to obtain

$$(12.36)\quad R_{\mu\nu} = \begin{Bmatrix} \alpha \\ \mu\ \alpha \end{Bmatrix}_{|\nu} - \begin{Bmatrix} \rho \\ \mu\ \nu \end{Bmatrix}_{|\rho} + \begin{Bmatrix} \alpha \\ \rho\ \nu \end{Bmatrix}\begin{Bmatrix} \rho \\ \mu\ \alpha \end{Bmatrix} - \begin{Bmatrix} \alpha \\ \rho\ \alpha \end{Bmatrix}\begin{Bmatrix} \rho \\ \mu\ \nu \end{Bmatrix}$$
$$= (\log\sqrt{-g})_{|\mu|\nu} - \begin{Bmatrix} \rho \\ \mu\ \nu \end{Bmatrix}_{|\rho} + \begin{Bmatrix} \alpha \\ \rho\ \nu \end{Bmatrix}\begin{Bmatrix} \rho \\ \mu\ \alpha \end{Bmatrix} - \begin{Bmatrix} \rho \\ \mu\ \nu \end{Bmatrix}(\log\sqrt{-g})_{|\rho}$$

Substitution of the various terms found in (12.32) to (12.35) into the expression (12.36) for $R_{\mu\nu}$ then yields

$$(12.37)\quad \begin{aligned} R_{00} &= \tfrac{3}{2}g'' + \tfrac{3}{4}g'^2 \\ R_{11} &= f'' + \frac{1}{r}f' - e^G(\tfrac{1}{2}g'' + \tfrac{3}{4}g'^2) \\ R_{22} &= r^2\left[\tfrac{1}{2}f'' + \tfrac{1}{4}f'^2 + \frac{3}{2}\frac{1}{r}f' - e^G(\tfrac{1}{2}g'' + \tfrac{3}{4}g'^2)\right] \\ R_{33} &= r^2\sin^2\theta\left[\tfrac{1}{2}f'' + \tfrac{1}{4}f'^2 + \frac{3}{2}\frac{1}{r}f' - e^G(\tfrac{1}{2}g'' + \tfrac{3}{4}g'^2)\right] \end{aligned}$$

By analogous calculations it is easily seen that $R_{\mu\nu} = 0$ if $\mu \neq \nu$. This was, of course, to be expected in view of the symmetries of the line element (12.29).

To obtain the mixed tensor $R^\mu{}_\nu$ we need only raise one index with $g^{\mu\nu}$, the inverse tensor to $g_{\mu\nu}$. From (12.29) these tensors are easily seen to be

$$(12.38)\quad g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -e^G & 0 & 0 \\ 0 & 0 & -e^G r^2 & 0 \\ 0 & 0 & 0 & -e^G r^2\sin^2\theta \end{pmatrix}$$
$$g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -e^{-G} & 0 & 0 \\ 0 & 0 & -\dfrac{e^{-G}}{r^2} & 0 \\ 0 & 0 & 0 & -\dfrac{e^{-G}}{r^2\sin^2\theta} \end{pmatrix}$$

Thus from (12.37) we obtain the diagonal mixed tensor $R^\mu{}_\nu$:

$$(12.39)\quad \begin{aligned} R^0{}_0 &= \tfrac{3}{2}g'' + \tfrac{3}{4}g'^2 \\ R^1{}_1 &= (\tfrac{1}{2}g'' + \tfrac{3}{4}g'^2) - e^{-G}\left(f'' + \frac{1}{r}f'\right) \\ R^2{}_2 = R^3{}_3 &= (\tfrac{1}{2}g'' + \tfrac{3}{4}g'^2) - e^{-G}\left(\tfrac{1}{2}f'' + \tfrac{1}{4}f'^2 + \frac{3}{2}\frac{1}{r}f'\right) \end{aligned}$$

Contraction of $R^\mu{}_\nu$ then leads to the scalar R:

$$(12.40)\quad R = 3(g'' + g'^2) - 2e^{-G}\left(f'' + \tfrac{1}{4}f'^2 + \frac{2}{r}f'\right)$$

In (12.39) and (12.40) we have all the quantities which enter the field equations. Thus we finally arrive at the explicit form of the Einstein field equations (12.4) in terms of f and g:

$$(12.41)\quad \begin{aligned} -\frac{8\pi\kappa}{c^2}T^0{}_0 &= \left[e^{-G}\left(f'' + \frac{f'^2}{4} + \frac{2f'}{r}\right) - \tfrac{3}{4}g'^2\right] + \Lambda \\ -\frac{8\pi\kappa}{c^2}T^1{}_1 &= \left[e^{-G}\left(\frac{f'^2}{4} + \frac{f'}{r}\right) - g'' - \tfrac{3}{4}g'^2\right] + \Lambda \\ -\frac{8\pi\kappa}{c^2}T^2{}_2 = -\frac{8\pi\kappa}{c^2}T^3{}_3 &= \left[e^{-G}\left(\frac{f''}{2} + \frac{f'}{2r}\right) - g'' - \tfrac{3}{4}g'^2\right] + \Lambda \\ -\frac{8\pi\kappa}{c^2}T^\mu{}_\nu &= 0 \qquad \text{for } \mu \neq \nu \end{aligned}$$

These will be useful in further simplifying the line element and also in explicitly solving the field equations, as we shall show in Chap. 13.

Let us discuss next the consequences of the condition of local isotropy on the form of the energy-momentum tensor $T^\mu{}_\nu$. We shall use a coordinate system in which the spatial line element is proportional to

$$d\sigma^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2$$

If we subject the neighborhood of the point considered to an orthogonal transformation of the space coordinates, the individual space components $T^m{}_n$ of the energy-momentum tensor should not change because of the assumed invariance of the energy-momentum distribution under such rotation. Since the matrix $T^m{}_n$ is invariant under orthogonal transformation, all its eigenvalues must be equal and it must be a multiple of the unit matrix; that is,

$$(12.42)\quad T^m{}_n = A\,\delta^m{}_n$$

where A is a scalar which may depend on x^0 and r. But since $\delta^m{}_n = g^m{}_n$, we have in (12.42) a tensor equation which holds in every coordinate system in three-dimensional space. In particular, if we use the polar coordinates r, θ, and φ, we also have

$$T^i{}_k = A g^i{}_k = A\,\delta^i{}_k \tag{12.43}$$

Thus, for every coordinate system,

$$T^1{}_1 = T^2{}_2 = T^3{}_3 \tag{12.44}$$

Inserting this into (12.41), we obtain the following relation on f:

$$f'' - \tfrac{1}{2}(f')^2 - \frac{1}{r} f' = 0 \tag{12.45}$$

This equation admits the first integral

$$f' = are^{f/2} \tag{12.46}$$

which leads to the general solution

$$e^f = \frac{b^2}{[1 - (ab/4)r^2]^2} \tag{12.47}$$

where a and b are arbitrary constants. When writing out the metric form, we can absorb the constant b^2 into the function $e^{g(x^0)}$ and define a new arbitrary constant by

$$|ab| = \frac{1}{r_0^2} \tag{12.48}$$

Then we obtain

$$ds^2 = (dx^0)^2 - e^{g(x^0)} \frac{1}{\left(1 + \dfrac{k}{4}\dfrac{r^2}{r_0^2}\right)^2} d\sigma^2 \tag{12.49}$$

where $k = 0, +1, -1$, corresponding to $ab = 0$, ab negative, and ab positive. These values describe a Euclidean space, a spherical space, and a pseudo-spherical space, as we shall presently discuss.

We can gain somewhat more insight into the nature of the Riemann space defined by the metric (12.49). From (12.41) and (12.44) it is evident that

$$G^1{}_1 = G^2{}_2 = G^3{}_3 = e^{-f}\left(\tfrac{1}{4}f'^2 + \frac{1}{r}f'\right)e^{-g} - g'' - \tfrac{3}{4}g'^2 \tag{12.50}$$

where we remind the reader that $g = g(x^0)$. On the other hand, a simple calculation with the explicit form (12.47) for e^f yields

$$e^{-f}\left(\tfrac{1}{4}f'^2 + \frac{1}{r}f'\right) = \frac{a}{b} = \text{const} \tag{12.51}$$

We find, therefore, not only that the space part of $G^\mu{}_\nu$ is isotropic,

$$G^1{}_1 = G^2{}_2 = G^3{}_3$$

but that it is constant throughout three-space at any given moment x^0. Indeed, from (12.50) and (12.51), we can say explicitly that

$$G^1{}_1 = G^2{}_2 = G^3{}_3 = e^{-g}\left(\frac{a}{b}\right) - g'' - \tfrac{3}{4}g'^2 \qquad g = g(x^0) \tag{12.52}$$

This demonstrates the complete homogeneity of three-space despite the apparent distinction of the center of the spherical coordinate system.

The form of line element (12.49) defines the so-called Robertson-Walker metric (Robertson, 1935). As we have shown above, it is characterized by the purely geometric property $G^1{}_1 = G^2{}_2 = G^3{}_3$. Thus, although we were guided by reasoning on the tensor $T^\mu{}_\nu$, we have arrived at a line element with a characteristic geometric structure which is interesting in its own right. It is therefore not surprising that the Robertson-Walker line element has been studied extensively in pure differential geometry theory. In the non-Euclidean geometries of the nineteenth century, the space elements of the Robertson-Walker metric played an important role. The case $k = -1$ occurred in the non-Euclidean geometry of Bolyai and Lobachevski, and the case $k = +1$ was first extensively discussed by Riemann. While the main interest of these mathematicians at that time centered around the problem of parallels in geometry, their basic geometric axioms led them to such high symmetry assumptions on the space considered that they were forced into the line element of (12.49). While the assumption on geometric isotropy of space is based on the principle of sufficient reason, that is, the argument that we do not know any reason why certain space directions should be distinguished from others, the isotropy of the distribution of matter is an observational fact. Thus, from the methodological point of view, Mach's principle plays a decisive role in establishing the Robertson-Walker line element in cosmology. The ultimate determining factor for geometry is the empirical law of matter distribution in the universe, as we have tried to show above.

12.4 Further Properties of the Robertson-Walker Metric

In order to study the dynamics of material particles in the Robertson-Walker metric, we need to determine their possible trajectories, i.e., the geodesics corresponding to (12.49). These are defined by

$$\frac{d^2x^\alpha}{ds^2} + \begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix} \frac{dx^\beta}{ds}\frac{dx^\gamma}{ds} = 0 \tag{12.53}$$

The most natural and important of these trajectories in the present case are those which describe a material particle with fixed space coordinates. As we stated in the preceding section, our distinguished three-space is characterized by the fact that in it matter is on the average at rest; thus the name co-moving coordinate system. It is therefore important that we verify that all curves $x^1 = \text{const}$, $x^2 = \text{const}$, $x^3 = \text{const}$ are indeed solutions of (12.52). It is evident that we need only show that the Christoffel symbol

$$\begin{Bmatrix} i \\ 0\ 0 \end{Bmatrix} = \frac{g^{i\rho}}{2}(g_{0\rho|0} + g_{\rho0|0} - g_{00|\rho}) \tag{12.54}$$

is zero; this, however, is immediately obvious from the form of the line element (12.49).

We have therefore verified that the Robertson-Walker metric represents a world in which a homogeneous distribution of matter is anchored to a co-moving coordinate system. On the other hand, a small test particle can certainly move along other geodesics, i.e., those for which ds need not be dx^0. Indeed, the theory of the initial-value problems for ordinary differential equations shows clearly that, if (12.52) has a solution such that all $\dot{x}^i$ vanish for a given moment x^0, we shall have $\dot{x}^i \equiv 0$ for all time. Conversely, if a test particle will have a nonzero velocity at one moment, it will never come to rest in the co-moving frame. Thus this frame may be considered to be a type of inertial frame. The everyday inertial and centrifugal forces which we experience on earth are precisely due to motion relative to this cosmic frame of reference, that of the so-called "fixed stars."

To determine completely the coordinate system, we also have to specify the spatial origin of the coordinates. Because of the equivalence of all points in three-space, this origin can be taken where most convenient,

depending on the particular problem considered, without affecting the form in which the line element is written. It will be specified in all forthcoming applications.

Next let us write the metric (12.49) in the more convenient form used by astronomers. In place of the radial distance r and the three-dimensional line element $d\sigma$, we define a dimensionless marker $u = r/r_0$ and a three-dimensional line element

$$d\chi^2 = \frac{1}{r_0^2}\,d\sigma^2 = du^2 + u^2(d\theta^2 + \sin^2\theta\, d\varphi^2) \tag{12.55}$$

Then, defining a new function $R^2(t) = r_0^2 e^{g(x^0)}$ with $t = x^0/c$, we can write

$$ds^2 = c^2\,dt^2 - R^2(t)\left[\frac{du^2 + u^2(d\theta^2 + \sin^2\theta\, d\varphi^2)}{[1 + (k/4)u^2]^2}\right] \tag{12.56}$$

Note that a given material point is considered to be labeled by co-moving coordinates u, θ, and φ, which are fixed in time, but that a physical distance interval between points, $R(t)\dfrac{d\chi}{1 + (k/4)u^2}$, is naturally time-dependent. However, the ratio of such a physical distance to $R(t)$, which we shall call the "radius" of the universe, is clearly time-independent, as one should expect.

In order to gain a more intuitive understanding of the metric form (12.56), let us digress somewhat for a moment to consider a problem in pure differential geometry. Suppose we have a four-dimensional Euclidean space with coordinates x^0, x^1, x^2, x^3 and a line element

$$ds^2 = (dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \tag{12.57}$$

If these coordinates are constrained to a hypersphere by the relation

$$(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 = R^2 \tag{12.58}$$

then clearly

$$x^0\,dx^0 + x^1\,dx^1 + x^2\,dx^2 + x^3\,dx^3 = 0 \tag{12.59}$$

This allows us to eliminate dx^0 from the line element (12.57) and obtain instead

$$ds^2 = \sum_{i=1}^{3} (dx^i)^2 + \frac{\left[d \sum_{i=1}^{3} (x^i)^2\right]^2}{4(x^0)^2} \tag{12.60}$$

In terms of the usual spherical coordinates ρ, θ, φ defined by

$$x^1 = \rho \cos \varphi \sin \theta \qquad x^2 = \rho \sin \varphi \sin \theta \qquad x^3 = \rho \cos \theta \tag{12.61}$$

this becomes, since $R^2 = \rho^2 + (x^0)^2$,

$$\begin{aligned} ds^2 &= d\rho^2 + \rho^2(d\theta^2 + \sin^2 \theta\, d\varphi^2) + \frac{\rho^2\, d\rho^2}{R^2 - \rho^2} \\ &= \frac{d\rho^2}{1 - \rho^2/R^2} + \rho^2(d\theta^2 + \sin^2 \theta\, d\varphi^2) \end{aligned} \tag{12.62}$$

With the convenient coordinate substitution

$$\rho = \frac{uR}{1 + \frac{1}{4}u^2} \tag{12.63}$$

Eq. (12.62) becomes

$$ds^2 = R^2 \left\{ \frac{du^2 + u^2(d\theta^2 + \sin^2 \theta\, d\varphi^2)}{(1 + \frac{1}{4}u^2)^2} \right\} \tag{12.64}$$

From the manner in which it has been constructed we may refer to this as the line element of a three-dimensional hypersphere of radius R which is imbedded in a four-dimensional Euclidean space.

It is now clear how the Robertson-Walker metric may be interpreted. Since (12.64) is identical with the space part of (12.56) for $k = 1$ we see that (12.56) describes an isotropic and homogeneous three-dimensional hypersphere with a uniform scalar curvature R given by

$$\frac{1}{R^2} = \frac{1}{R^2(t)} \tag{12.65}$$

This explains the name "radius of the universe," which we used for $R(t)$ (for $k \neq 1$, see Exercise 12.1). One must be careful to note that this geometric picture is without any observable consequences. The imbedding four-dimensional Euclidean space is entirely fictitious and serves only to enhance our intuitive understanding of the Robertson-Walker metric.

In the case $k = 1$, the hypersphere, the Robertson-Walker metric has some elegant geometrical properties worthy of further comment. To discuss these we introduce as a new radial coordinate the hyperspherical angle y, defined by

$$\sin y = \frac{u}{1 + u^2/4} \qquad dy^2 = \frac{du^2}{(1 + u^2/4)^2} \tag{12.66}$$

Since the range of u is 0 to ∞, the range of y is 0 to π. The Robertson-Walker metric in terms of y is

$$ds^2 = c^2\, dt^2 - R^2(t)[dy^2 + \sin^2 y(\sin^2 \theta\, d\varphi^2 + d\theta^2)] \tag{12.67}$$

Let us now calculate the spatial distance from the origin of the coordinate system to the farthest point along a ray, $d\varphi = d\theta = 0$. This will be simply

$$\int_0^\pi R\, dy = \pi R \tag{12.68}$$

This may be identified as half the circumference of the hypersphere of radius R, an elegant result. We may similarly calculate the volume of a part of three-space at a given time; the invariant volume element of three-space will be

$$\sqrt{|g|}\, dy\, d\theta\, d\varphi = R^3 \sin^2 y\, dy \sin \theta\, d\theta\, d\varphi$$

as discussed in Sec. 3.2. Thus a sphere extending from $y = 0$ to y_0 has a volume given by

$$V(y_0) = 4\pi R^3 \int_0^{y_0} \sin^2 y\, dy = 4\pi R^3 \frac{2y_0 - \sin 2y_0}{4} \tag{12.69}$$

For very small y_0 this is $4\pi(Ry_0)^3/3$, which allows us to identify Ry_0 as a convenient measure of small radial distances. The total volume of the three-space is given for $y_0 = \pi$ as

$$V = 2\pi^2 R^3 \tag{12.70}$$

Lastly let us calculate the area of a sphere of coordinate radius y_0. Analogous to the volume element above, the surface area element will be $R^2 \sin^2 y_0 \sin^2 \theta\, d\theta\, d\varphi$. Thus the total area will be

That is, analogous to the marker r in the Schwarzschild metric, the *coefficient* $R^2 \sin^2 y$ of the angular interval $d\theta^2 + \sin^2 \theta \, d\varphi^2$ plays a distinguished role as a radial marker in terms of physical area; its square times 4π is the area of a sphere. (See Exercise 12.5 for interesting features of A.)

The cases of $k = 0$ and $k = -1$ may be thought of in geometrical terms also: $k = 0$ represents a flat Euclidean space with a time-dependent scale factor $R(t)$, while $k = -1$ represents a hypersurface of constant negative curvature, roughly analogous to the shape of the top of a saddle in two dimensions (see Exercises 12.1 to 12.4). Unfortunately a two-surface of constant negative curvature cannot be imbedded in Euclidean three-space, as recognized by Bolyai and Lobachevski, thus making it difficult to visualize, and similarly for a hypersurface of constant negative curvature.

In our derivation of the Robertson-Walker metric we postulated the existence of a universal time-coordinate and a Gaussian coordinate system in the large. Gödel (1949) obtained a cosmological solution of Einstein's equations which does not satisfy this postulate, but which does correspond to a constant matter density. There are difficulties associated with the physical interpretation of this solution, however. For example, the line element contains a cross term of the form $dx^0 \, dx^2$ and corresponds to an intrinsically rotating universe. We therefore defer discussion of this solution until the next chapter.

12.5 The Red Shift and the Robertson-Walker Metric: Hubble's Law

We shall now show that the Robertson-Walker metric (12.56) gives rise to an apparent shift in frequency of light emitted by distant objects, in agreement with the observed Hubble law. We consider a radiating object, say a galaxy G_e which is considered as a particle in our model. We then suppose that its light is observed by an observer in galaxy G_0 placed at the origin of the co-moving coordinates. The galaxy G_e is characterized by its distance marker u or, equivalently, by another useful *time-independent* marker l defined by the relation

$$dl = \frac{d\chi}{1 + (k/4)u^2} = \frac{[du^2 + u^2(d\theta^2 + \sin^2 \theta \, d\varphi^2)]^{1/2}}{1 + (k/4)u^2} \tag{12.72}$$

In terms of the universal time used in the Robertson-Walker metric, the light emitted by G_e at time t_e is observed on G_0 at a time t_0 with $t_0 > t_e$.

Since light travels along a null geodesic, we must have $ds^2 = 0$, or

$$\int_{t_e}^{t_0} \frac{c\,dt}{R(t)} = l \tag{12.73}$$

which relates the distance marker l to the time difference $t_0 - t_e$.

Consider now the light emitted by G_e at time $t_e + \Delta t_e$. It will be received by G_0 at time $t_0 + \Delta t_0$, where Δt_0 will be determined by the relation

$$\int_{t_e+\Delta t_e}^{t_0+\Delta t_0} \frac{c\,dt}{R(t)} = l \tag{12.74}$$

since l is a fixed marker distance in the co-moving coordinates. Consider Δt_e as the period of some periodic physical phenomenon taking place on G_e, the emission of radiation for instance, and Δt to be short compared with the travel time from G_e to G_0. The periodic phenomenon will appear, as seen from G_0, to have a period Δt_0, which, from (12.73) and (12.74), will be such that the increment of the l integral is zero; thus, by elementary calculus,

$$\frac{\Delta t_0}{R(t_0)} - \frac{\Delta t_e}{R(t_e)} = 0 \tag{12.75}$$

Calling $R(t_0) = R_0$ and $R(t_e) = R_e$ and writing (12.75) in terms of frequencies, we obtain

$$\frac{\nu_e}{\nu_0} = \frac{R_0}{R_e} \tag{12.76}$$

For radiation propagating with velocity c, we associate a wavelength λ with a frequency ν, which, in view of the defining relation $c = \lambda\nu$, gives

$$\frac{\lambda_0}{\lambda_e} = \frac{R_0}{R_e} \tag{12.77}$$

From this we define a relative shift in wavelength with respect to the wavelength at emission:

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{R_0}{R_e} - 1 \tag{12.78}$$

From the above it is clear that the radiation emitted at one point will appear with a shift in wavelength at another point. This phenomenon discussed in Sec. 12.1. That is, λ_0 at reception is larger than λ_e at emission. Therefore, from (12.77), the function $R(t)$ must at present be monotonically increasing with time since the time of reception t_0 is later than the time of emission t_e. We thus must have a universe in expansion.

Hubble's law (12.2) asserts a linear relation between the red shift z and distance of a galaxy. We shall now derive this from (12.73) and (12.78) in first approximation. We begin by introducing a power-series expansion in terms of the travel time of light. Observe that $[R(t)]/c$ has the dimension of time and can be interpreted as the time needed by a light ray to reach a distance of the magnitude of the radius of the universe. We shall consider physical phenomena which require considerably smaller time intervals and use power-series expansions in terms of the quantity $[c(t_0 - t_e)]/R_0$, which can be considered as small. Formula (12.73) then tells us that l will be smaller than 1.

Expanding $1/[R(t)]$, we obtain

$$\frac{1}{R(t)} = \frac{1}{R_0} - \frac{R_0'}{R_0 c}\frac{(t-t_0)c}{R_0} + \frac{1}{c^2}\left[\frac{(R_0')^2}{R_0} - \frac{R_0''}{2}\right]\left[\frac{(t-t_0)c}{R_0}\right]^2 + O\left[\frac{c(t-t_0)}{R_0}\right]^3 \tag{12.79}$$

where the prime denotes differentiation with respect to t. Next, expanding (12.73) and (12.78) in powers of $[c(t_0 - t_e)]/R_0 = h$, we obtain

$$l = h + \frac{1}{2}\frac{R_0'}{c}h^2 + O(h^3) \tag{12.80}$$

and

$$z = \frac{R_0}{R_e} - 1 = \frac{R_0'}{c}h + \frac{R_0}{c^2}\left[\frac{(R_0')^2}{R_0} - \frac{R_0''}{2}\right]h^2 + O(h^3) \tag{12.81}$$

Eliminating h between the two equations above, we get

$$\begin{aligned} cz &= R_0'l + \frac{1}{2c}(R_0'^2 - R_0''R_0)l^2 + O(l^3) \\ &= R_0'l + \frac{R_0'^2l^2}{2c}(1 + q_0) \end{aligned} \tag{12.82}$$

where

$$q_0 = -\frac{R_0''R_0}{R_0'^2} \tag{12.82$'$}$$

is referred to as the deceleration parameter because of its linear relation to $-R_0''$. This formula, which relates the red shift to the marker distance, will be of fundamental importance in establishing a theoretical relation between red shift and astronomical distance, both observable quantities to the astronomer.

Let us consider for the time being only the lowest-order term in (12.82) and remember that the physical distance which an astronomer measures is, in first approximation, $L \cong R_0 l$. Formula (12.82) then becomes

$$cz \cong \frac{R_0'}{R_0}(R_0 l) \cong \frac{R_0'}{R_0}L \tag{12.83}$$

This is precisely Hubble's red shift law (11.2) if we identify H with R_0'/R_0.

Hubble's law can also be interpreted as a Doppler effect on the basis of the Robertson-Walker metric (12.56). The galaxies G_0 and G_e have fixed co-moving coordinates, but the physical distance between them, $L = R(t)l$, increases with time. Indeed, the velocity is clearly

$$v = R'(t)l = \frac{R'(t)}{R(t)}L \tag{12.84}$$

Thus Doppler's law gives

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c} = \frac{R_0'L}{R_0 c} \tag{12.85}$$

and we again obtain Hubble's law (12.2).

In conclusion we can say that *the Robertson-Walker metric explains the red shift* and allows us to identify Hubble's "constant" in first approximation as

$$H = \frac{R_0'}{R_0} \tag{12.86}$$

Further considerations on the red shift will be taken up in the next section.

12.6 The Apparent Magnitude–Red Shift Relation

In the preceding section we considered the red shift as a function of the distance of an emitting galaxy and obtained the Hubble law as a first approximation. Unfortunately the measurement of the distance of a galaxy is an indirect and uncertain process, as we have already indicated. In contrast, we now consider the two parameters of a galaxy most easily measured, its red shift z and its apparent luminosity e_0, which is the amount of energy received from the star per unit time and unit area. We can relate these directly measurable parameters using only our knowledge of the geometry of space embodied in the Robertson-Walker metric; this section will indicate how this is done.

Let us use the usual polar coordinate system (u,θ,φ) and place the origin on the emitting galaxy G_e. An implicit relation exists between the red shift z and the marker u_0 of the observer's galaxy, which can in principle be obtained from (12.72), (12.73), and (12.78): from (12.73) and (12.72) we can relate u_0 to $R(t)$, and (12.78) relates z to $R(t)$. We may write this implicit relation symbolically as

$$A(z,u_0) = 0 \tag{12.87}$$

The next step of the program is to replace u_0 by the physically observable quantity e_0. Our knowledge of the Robertson-Walker metric allows us to do this on purely geometrical grounds if the total rate of emission E_e of the galaxy is known. We shall show that one can indeed obtain a relation of the form

$$B\left(\frac{e_0}{E_e}, u_0\right) = 0 \tag{12.88}$$

By elimination of u_0 between relations A and B we thus arrive at the desired relation, in symbolic form,

$$C\left(\frac{e_0}{E_e}, u_0\right) = 0 \tag{12.89}$$

This links the observable parameters e_0 and z if E_e is known. As we have discussed, astronomers know how to estimate E_e for some specific types of galaxies, so that (12.89) may be applied to a group of galaxies of similar type and therefore with roughly the same E_e.

Let us now carry out the above indicated operations and obtain an explicit and useful form of the apparent magnitude–red shift relation $C(e_0/E_e, z) = 0$. We need, first, to review a few elementary notions and definitions of astronomy. One defines a logarithmic scale to measure luminosities by defining the *apparent magnitude* of a star as

$$m_0 = -2.5 \log_{10} e_0 + \text{const} \qquad \text{(log to base 10)} \tag{12.90}$$

The zero of this logarithmic scale is fixed arbitrarily by defining the apparent magnitude of the North Star to be $m_0 = 2.15$. For an arbitrary star, one then has

$$m_0 = -2.5 \log_{10} \frac{e_0}{e_{\text{N.S.}}} + 2.15 \tag{12.91}$$

where $e_{\text{N.S.}}$ is the energy flux from the North Star. Such a procedure for measuring the brightness of a star finds its origin in the measurements made with the naked eye by early astronomers, who used an additive scale. It is now known that the response of the eye varies as the logarithm of the exciting intensity; hence the definition (12.91) of the apparent magnitude and its normalization to make the numbers agree with old catalogues of magnitudes. The luminosity of stars and galaxies is now measured more directly by means of instruments such as the bolometer, which gives e_0 directly. In practice one measures the red shift and the apparent magnitude of whole galaxies, and not of individual stars, for it is only for distant galaxies that the red shift due to the general recession is large enough to be measured and distinguished from the Doppler shift due to random velocities which are always present.

In order to calculate $C(e_0/E_e, z)$ explicitly, let us first derive the relation $B(e_0/E_e, u_0) = 0$ from the general form of the Robertson-Walker metric. Consider the total energy received per unit time on the whole surface of a sphere reached by light in a time $(t_0 - t_e)$. This energy is smaller than E_e for two reasons, which are best put in evidence by using the photon picture of light: First, the photons received have a degraded energy due to reddening by a factor $\nu_0/\nu_e = R_e/R_0$, where $R_e = R(t_e)$ and $R_0 = R(t_0)$. Second, if we consider the photons emitted at regular intervals Δt_e, they will arrive separated by longer intervals Δt_0, and the two rates will be in the ratio $\Delta t_e/\Delta t_0 = R_e/R_0$. Thus the energy received per unit time on the whole sphere surrounding the galaxy at a coordinate distance u_0 will be

$$E_0 = E_e\left(\frac{R_e}{R_0}\right)^2 \tag{12.92}$$

From the form of the Robertson-Walker metric in polar coordinates,

$$ds^2 = (dx^0)^2 - \frac{R(t)^2}{[1 + (k/4)u^2]^2}(du^2 + u^2\,d\theta^2 + u^2 \sin^2\theta\,d\varphi^2) \tag{12.93}$$

it is evident that a solid angle defined by $d\theta$ and $d\varphi$ corresponds at a coordinate distance u_0 to a physical area of

$$dS = \left(\frac{R_0 u_0}{1 + (k/4)u_0^2}\right) d\theta \left(\frac{R_0 u_0 \sin\theta}{1 + (k/4)u_0^2}\right) d\varphi \qquad R_0 = R(t_0) \tag{12.94}$$

Thus the total physical area of a sphere of coordinate radius u_0 is

$$S = \frac{4\pi R_0^2 u_0^2}{[1 + (k/4)u_0^2]^2} \tag{12.95}$$

The total energy received per unit area and unit time interval on the earth considered as a point on this sphere is therefore given by

$$e_0 = \frac{E_0}{S} = E_e \left(\frac{R_e}{R_0}\right)^2 \frac{[1 + (k/4)u_0^2]^2}{4\pi R_0^2 u_0^2} \tag{12.96}$$

This is precisely the quantity measured by astronomers using a bolometer. The apparent magnitude as defined by (12.91) thus becomes

$$m_0 = 5 \log_{10} \left[\frac{R_0 u_0}{R_e\left(1 + \frac{k}{4}u_0^2\right)}\right] - 2.5 \log_{10} E_e - \text{const} \tag{12.97}$$

Since the light coming from all stars is received at the same time of observation t_0, we have absorbed some $\log_{10} R_0$ factors in the constant in (12.97).

The ratio R_0/R_e appearing in (12.97) can be immediately expressed as a function of z through (12.78):

$$1 + z = \frac{R_0}{R_e} \tag{12.98}$$

This equation also gives in principle $t_0 - t_e$ as a function of z. From the fact that $ds^2 = 0$ along the path of a light ray, we can also obtain the expression $u_0/[1 + (k/4)u_0^2]$ as a function of t_e from (12.72) and (12.73):

$$\int_{t_e}^{t_0} \frac{c\,dt}{R(t)} = \int_0^{u_0} \frac{du}{1 + (k/4)u^2} \tag{12.99}$$

and then as a function of z, using (12.98). Therefore, once a specific function $R(t)$ is given, m_0 can be expressed as a function of z alone by using (12.99). That is, we may write

$$m_0 = 5 \log_{10} F(z) - 2.5 \log_{10} E_e + \text{const} \tag{12.100}$$

To give a simple and explicit illustration of the preceding paragraph, consider the case of an expanding Euclidean three-space, $k = 0$, and assume we are dealing with a model in which Hubble's law is strictly valid in the form

$$cz = HL \tag{12.101}$$

(see Sec. 13.6 also). Here L is the astronomical distance marker defined at the beginning of this chapter by the law of energy decrease with the inverse square of the distance. In this case, from the definition of L and the fact that the two-space is Euclidean, (12.96) can be replaced by

$$e_0 = \frac{E_e}{4\pi L^2} = \frac{E_e}{4\pi} \frac{H^2}{(cz)^2} \tag{12.102}$$

Therefore, by the use of (12.102), we see that (12.91) becomes, in this special case,

$$m_0 = 5 \log_{10} (cz) - 2.5 \log_{10} E_e + \text{const} \tag{12.103}$$

Indeed, this formula is the exact translation of Hubble's law in terms of the variables m_0 and z.

As a second example of how an explicit apparent magnitude–red shift relation is obtained, we shall consider the completely general relation (12.97) and assume that the quantities z, l, and u_0 are small. It is then possible to obtain a series expansion of m_0 in terms of z for an arbitrary function $R(t)$ (Robertson, 1955; Hoyle and Sandage, 1956). The calculation of the first two terms of m_0 is straightforward and requires that we retain only second-order terms in z and l throughout. Let us begin by writing (12.97) as

$$m_0 = 5 \log_{10} \frac{R_0}{R_e} + 5 \log_{10} \frac{u_0}{1 + (k/4)u_0^2} + \text{const} \tag{12.104}$$

where the constant depends upon the type of galaxy considered. Using (12.78), the first term is readily rewritten in terms of z, and we obtain

$$m_0 = 5 \log_{10} (1 + z) + 5 \log_{10} \frac{u_0}{1 + (k/4)u_0^2} + \text{const} \tag{12.105}$$

To express the second term as a function of z, first note that (12.72) can be easily integrated to give

$$l = \int_0^{u_0} \frac{du}{1 + (k/4)u^2} = \frac{2}{\sqrt{k}} \arctan \left(\frac{\sqrt{k}}{2} u_0 \right) \tag{12.106}$$

and thus

$$u_0 = \frac{2}{\sqrt{k}} \tan \left(\frac{\sqrt{k}}{2} l \right) \tag{12.107}$$

An elementary calculation then yields the result that the argument of the second logarithm in (12.105) is

$$\frac{u_0}{1 + (k/4)u_0^2} = \frac{1}{\sqrt{k}} \sin (\sqrt{k}\, l) = l + O(l^3) \tag{12.108}$$

It thus only remains for us to express l as a series in z. To do this we invert Eq. (12.82) to obtain

$$l = \frac{cz}{R_0'} \left[1 - \frac{1}{2} \left(\frac{1 + q_0}{R_0'} \right) z + \cdots \right] \tag{12.109}$$

We can thus express m_0 explicitly in terms of z by substituting (12.109) and (12.108) into (12.105):

$$\begin{aligned} m_0 &= 5 \log_{10} (1 + z) + 5 \log_{10} cz \\ &\quad + 5 \log_{10} \left[1 - \frac{1}{2} \left(\frac{1 + q_0}{R_0'} \right) z + \cdots \right] + \text{const} \end{aligned} \tag{12.110}$$

where we have absorbed a new term, $-5 \log_{10} R_0'$, into the constant. This expression is easily expanded in terms of z; one must only be careful to remember that the logarithms are to base 10. The final result to first order in z is

$$\begin{aligned} m_0 &= 5 \log_{10} cz + \frac{5}{2 \log_e 10} (1 - q_0)z + \text{const} \\ &= 5 \log_{10} cz + (1.086)(1 - q_0)z + \text{const} \end{aligned} \tag{12.111}$$

which is completely general and holds for any function $R(t)$ having a reasonable enough behavior so that the power series we have used are valid.

One should notice that, to lowest order in z, the result (12.111) is identical with the previous simple relation (12.103).

Because of the logarithmic relation of m_0 to e_0 a plot of m_0 versus z (12.111) clearly does not give an observational value of H but does test the functional form of the Hubble law and yields a value for the deceleration parameter. Using the brightest galaxies in clusters, correcting for a variation of luminosity as a function of frequency, and allowing for a change in luminosity as a function of time, astronomers now estimate that (Sandage, 1972a, 1972b, 1972c)

$$q_0 = 1.0 \pm 0.5 \tag{12.112}$$

This value remains in doubt due to uncertainties in the corrections mentioned above, but it appears likely that q_0 is positive; from the definition (12.82′) this would mean that the universe is decelerating, $R_0'' < 0$. We can hope that the value of q_0 will eventually be determined to better accuracy with the use of radio galaxies, and perhaps also with the use of quasars if a classification scheme can be obtained to correlate absolute luminosities.

We now see that astronomical measurements are able to supply us with values for the Hubble constant H, the deceleration parameter q_0, and the density of material in the universe. In the following chapter on cosmological models we shall see how a knowledge of these parameters can be used to choose among various models and moreover how the Einstein equations can be used to relate q_0 and ρ_0 to k, thereby determining in principle the shape of the universe.

Exercises

12.1 In place of (12.63) we may introduce a coordinate substitution that formally encompasses the cases $k = 0$ and $k = -1$, as well as $k = 1$ as discussed in the text. Analogous to (12.63) we introduce

$$\rho = \frac{\sqrt{k}\, Ru}{1 + (k/4)u^2} \qquad \sqrt{k}\, R = \tilde{R}$$

where $\tilde{R}$ is necessarily finite and real. Thus for $k = 1$ R must be real, for $k = -1$ it must be imaginary, and for $k = 0$ it must be infinite in the limiting sense that $\tilde{R}$ remains finite as $k \to 0$. Show that the three-space metric is

$$ds^2 = \tilde{R}^2 \left[\frac{du^2 + u^2(d\theta^2 + \sin^2\theta\, d\varphi^2)}{[1 + (k/4)u^2]^2}\right]$$

analogous to (12.64).

12.2 (*continued*) The $k = 1$ case can be "visualized" as a three-dimensional hypersphere imbedded in a fictitious four-dimensional space, as we showed in the text. How can the $k = 0$ case be visualized? Can the $k = -1$ case be visualized in an analogous way to $k = 1$? (This corresponds to the classical non-Euclidean geometry studied by Bolyai and Lobachevski. The three-hypersurface is referred to as a pseudo-hypersphere.)

12.3 Give a derivation of the Robertson-Walker metric from purely geometrical considerations based on plausibility arguments; use the three basic demands of Sec. 12.3. Do this by asking what two-dimensional surfaces are homogeneous and isotropic, then extend the question to three dimensions by analogy.

12.4 Analyze the $k = -1$, or pseudo-hyperspherical, universe in a similar manner to the hyperspherical case by introducing an angle defined by

$$\sinh y = \frac{u}{1 - u^2/4}$$

What are the analogues of the circumference of the universe (12.68), the volume expression (12.69), and the total volume of the hyperspherical universe (12.69′)?

12.5 From (12.71) plot A as a function of y_0. Note that it is maximum for $y_0 = \pi/2$, a point one-fourth of the way around the universe, and becomes zero for $y_0 = \pi$, the largest sphere that can be constructed. Give a physical or geometrical interpretation of this.

12.6 What is the Petrov type of a space-time with a Robertson-Walker metric for the three possible values of k? (See Exercise 10.8.)

Problems

12.1 In principle it is possible to make interferometric measurements of a star's position which are extremely accurate if the base line of the interferometer is sufficiently large. Show that the distance to which such methods could be used with a radio interferometer equal in size to the earth's orbital radius is of the order of 10^9 light years. (The earth's orbit is about 500 light seconds in radius and the wavelength of radio waves used could be of order 1 cm.) Such a system might be devised in practice using, for example, a satellite in solar orbit as one end of the interferometer and the earth as the other (see Weinberg, 1971, p. 429).

12.2 (*continued*) Discuss how the satellite interferometer system might work. Would it seem feasible to devise an optical interferometer of the scale of the solar system?

Bibliography

Alpher, R. A. (1948): A Neutron-Capture Theory of the Formation and Relative Abundance of the Elements, *Phys. Rev.*, **74**:1577.

Beer, A. (ed.) (1960): "Vistas in Astronomy," vol. 3, New York, p. 320.

Burbidge, G., and M. Burbidge (1967): "Quasi Stellar Objects," San Francisco.

Dicke, R. H. (1970): "Gravitation and the Universe," lecture III, American Philosophical Society, Philadelphia.

Dicke, R. H., P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson (1965): Cosmic Black Body Radiation, *Astrophys. J.*, **142**:414.

Einstein, A. (1918): Prinzipielles zur allgemeinen Relativitätstheorie, *Ann. Physik*, **55**(4):241–244.

Gamow, G. (1946): Expanding Universe and the Origin of the Elements, *Phys. Rev.*, **70**:572.

Gödel, K. (1949): An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation, *Rev. Mod. Phys.*, **21**:447–450.

Hoyle, F., and A. Sandage (1956): The Second Order Term in the Redshift-Magnitude Relation, *Pub. Astron. Soc. Pacific*, **68**:301.

Hubble, E. P. (1936): The Luminosity Function of Nebulae, II, *Astrophys. J.*, **84**:270–295.

Hubble, E. P. (1957): The Exploration of Space, in M. Munitz (ed.), "Theories of the Universe," Glencoe, Ill.

Leavitt, H. S. (1912): Periods of 25 Variable Stars in the Small Magellanic Cloud, *Harvard Circ.* 173.

Lichnerowicz, A. (1955): "Théories relativistes de la gravitation et de l'électromagnétisme," Paris.

Mach, E. (1883): "Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt," Leipzig.

Morrison, P. (1973): Resolving the Mystery of the Quasars, *Physics Today*, **26**:23.

Oke, J. G. (1959): The Hertzsprung-Russel Diagram for F5-K2 Stars with the Most Accurate Absolute Magnitudes, *Astrophys. J.*, **130**:487–495.

Oort, J. H. (1958): Distribution of Galaxies and the Density in the Universe, in "La Structure et l'évolution de l'univers," Institut International de Physique Solvay, Brussels.

Patterson, C. C. (1960): Age of the Meteorites and the Earth, *Geochim. Cosmochim. Acta*, **10**:280.

Peebles, P. J. E. (1967): Microwave Radiation from the Big Bang in J. Ehlers (ed.), "Relativity Theory and Astrophysics, 1: Relativity and Cosmology," vol. 8 of Lectures in Applied Mathematics, Providence.
Penzias, A. A., and R. W. Wilson (1965): A Measurement of Excess Antenna Temperature at 4080 Mc/s, *Astrophys. J.*, **142**:419.
Pontecorvo, B., and J. Smorodinsky (1962): The Neutrino and the Matter Density in the Universe, *J. Expt. Theoret. Phys.* (transl.), **14**:173–176.
Robertson, H. P. (1935): Kinematics and World Structure, *Astrophys. J.*, **82**:284–301.
Sandage, A. (1958): Current Problems in the Extragalactic Distance Scale, *Astrophys. J.*, **127**:513–526.
Sandage, A. (1961): The Ability of the 200 Inch Telescope to Discriminate between Selected World Models, *Astrophys. J.*, **133**:355.
Sandage, A. (1972): Distance to Galaxies: The Hubble Constant, the Friedmann Time, and the Edge of the World, *Proc. Symp. Galaxy Distance Scale, Essex, England.*
Sandage, A. (1972a): The Redshift-Distance Relation, I: Angular Diameter of First-ranked Cluster Galaxies as a Function of Red Shift: The Aperture Correction to Magnitude, *Astrophys. J.*, **173**:485.
Sandage, A. (1972b): The Redshift-Distance Relation, II: The Hubble Diagram and Its Scatter for First-ranked Cluster Galaxies: A Formal Value for q, *Astrophys. J.*, **178**:1.
Sandage, A. (1972c): The Redshift-Distance Relation, III: Photometry and the Hubble Diagram for Radio Source and the Possible Turn-on time for QSOs, *Astrophys. J.*, **178**:25.
Schmidt, M. (1967): Lectures on Quasi Stellar Objects, in J. Ehlers (ed.), "Relativity Theory and Astrophysics, 1: Relativity and Cosmology," Providence.
Weinberg, S. (1972): "Gravitation and Cosmology," New York, chap. 14.
Wilson, O. C. (1950): A Colour Magnitude Diagram for Late Type Stars near the Sun, *Astrophys. J.*, **130**:496–498.

CHAPTER

13

Cosmological Models

So far our considerations have been of a purely geometrical and rather general nature; that is, we have dealt with the form of $g_{\mu\nu}$ alone. There remains considerable latitude in the functional form of $R(t)$ and the values of the constants Λ and k which appear in the Robertson-Walker metric (12.56). We begin this chapter by pursuing the consequences of various specifications of these quantities. The consideration of these "normal" models will be followed by two brief sections on more "heretical" models: the steady-state model, which does not obey Einstein's equations, and the Gödel model, which does not have a Robertson-Walker type of metric. We conclude by discussing the converse of the apparent magnitude–red shift problem.

13.1 Einstein's Equations and the Robertson-Walker Metric

In the following sections we shall consider a fluid continuum of a highly idealized nature which consists of galactic clusters. This fluid will be described by an average density ρ and an average internal pressure p, both of which can be functions of time but not of space. From this we are led to a very simple general form for $T^\mu{}_\nu$. The co-moving coordinates of a galactic cluster on the average satisfy $\dot{x}^0 = 1$, $\dot{x}^1 = \dot{x}^2 = \dot{x}^3 = 0$, so from the fluid energy-momentum tensor (10.41) and the fact that $g_{00} = 1$ in the Robertson-Walker metric (12.56), we obtain

$$T^\mu{}_\nu = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -\dfrac{p}{c^2} & 0 & 0 \\ 0 & 0 & -\dfrac{p}{c^2} & 0 \\ 0 & 0 & 0 & -\dfrac{p}{c^2} \end{pmatrix} \tag{13.1}$$

Note that this expression is clearly consistent with the Robertson-Walker metric since $T^1{}_1 = T^2{}_2 = T^3{}_3$ and $T^\mu{}_\nu = 0$ for $\mu \neq \nu$. The relative sizes of ρ and p/c^2 can be estimated by assuming, as is usual, that the pressure is due to the residual random motion of galactic clusters, that is, to local deviations from the average state. It is a well-known result of the kinetic theory of gases that the relation between pressure and density is $p/\rho = \hat{v}^2/3$, where $\hat{v}$ is the root-mean-square random velocity of the gas molecules. Applying this to the galactic-cluster gas of the universe, we see that

$$\frac{p/c^2}{\rho} = \frac{1}{3}\frac{\hat{v}^2}{c^2} \tag{13.2}$$

The observed random motions of galaxies are in general much less than c, so that, in most reasonable models of the universe, we expect the pressure component at the present epoch in the evolution of the universe.

From a comparison of the Robertson-Walker metric in the two forms (12.29) and (12.56) we see that

$$e^{G(x^0,r)} = \frac{R(t)^2}{r_0^2(1 + kr^2/4r_0^2)^2} \tag{13.3}$$

$$e^{g(x^0)} = R(t)^2 \qquad e^{f(r)} = \frac{1}{r_0^2(1 + kr^2/4r_0^2)^2}$$

Using $T^\mu{}_\nu$ in (13.1) and $G^\mu{}_\nu$ as obtained from a short calculation with (13.3) and (12.41), we can write the Einstein equations (12.4) in the form

$$\frac{8\pi\kappa}{c^2}\rho = -\Lambda - G^0{}_0 = -\Lambda + \left[\frac{3k}{R(t)^2} + \frac{3R'(t)^2}{c^2R(t)^2}\right] \tag{13.4a}$$

$$\frac{8\pi\kappa}{c^2}\left(\frac{p}{c^2}\right) = \Lambda + G^i{}_i = \Lambda - \left[\frac{k}{R(t)^2} + \frac{R'(t)^2}{c^2R(t)^2} + \frac{2R''(t)}{c^2R(t)}\right] \tag{13.4b}$$

For later use we shall now also obtain an equivalent system. By taking linear combinations we get two new equations with different structure; one contains only a second-order derivative, and the other is independent of Λ:

$$\frac{4\pi\kappa}{c^2}\left(\rho + \frac{3p}{c^2}\right) = \Lambda - \frac{3R''(t)}{c^2R(t)} \tag{13.5a}$$

$$\frac{4\pi\kappa}{c^2}\left(\rho + \frac{p}{c^2}\right) = \frac{k}{R^2(t)} + \frac{R'^2(t) - R(t)R''(t)}{c^2R^2(t)} \tag{13.5b}$$

We can already draw an important conclusion from these equations if we combine (13.4a) with (13.5b). We may bring (13.5b) into the form

$$\frac{d}{dt}\left[\frac{1}{c^2}\frac{R'(t)}{R(t)}\right] = \frac{k}{R(t)^2} - \frac{4\pi\kappa}{c^2}\left(\rho + \frac{p}{c^2}\right) \tag{13.6}$$

On the other hand, by differentiating (13.4a) with respect to time, we obtain

$$\frac{8\pi\kappa}{c^2}\frac{d\rho}{dt} = -\frac{6k}{R^3}R' + \frac{6R'}{R}\frac{d}{dt}\left(\frac{1}{c^2}\frac{R'}{R}\right) \tag{13.7}$$

Inserting (13.6) into (13.7) and rearranging, we obtain the differential identity

$$\frac{d}{dt}(\rho R^3) + \frac{p}{c^2}\frac{dR^3}{dt} = 0 \tag{13.8}$$

Let us take an element of volume in the co-moving coordinates. Its geometric content will be proportional to R^3, from the Robertson-Walker metric. Thus, if $V(t)$ denotes the actual volume content of the element considered, $M = \rho V$ would measure the mass content and (13.8) could be written as

$$dM + \frac{1}{c^2}p\,dV = 0 \tag{13.9}$$

Since c^2M measures the energy content of the element and $p\,dV$ the work done against the pressure forces, we see that the energy balance under the cosmic evolution is preserved. We may say equivalently that the entropy change under the evolution is zero; see Exercise 13.1.

If we prescribe the equation of state of the fluid, $p = p(\rho)$, we can obtain from (13.8) the dependence of R upon ρ. Indeed, we may write Eq. (13.8) in the alternative form

$$\frac{dR}{R} + \frac{1}{3}\left[\frac{d\rho}{\rho + p(\rho)/c^2}\right] = 0 \tag{13.10}$$

From this, the relation between R and ρ is obtained by simple integration. In the case of a pressure-free model, we find

$$R^3\rho = \text{const} \tag{13.11}$$

and in the case of the ideal gas law, $p = \alpha\rho$, we obtain

$$R^{3(1+\alpha/c^2)}\rho = \text{const} \tag{13.12}$$

In particular, if the pressure is entirely due to radiation, one has the well-known law $p = (c^2/3)\rho$ and the result $R^4\rho = \text{const}$.

At this point we would do well to note that ρ is only a *parameter*, which we may interpret as a *coordinate density* and not necessarily a physical density. Thus we cannot in general say that the mass M of a body is given in terms of the physical volume V by $M = \rho V$, except in the classical limit or in Euclidean space. In general the gravitational binding energy will contribute to M. This will be demonstrated and discussed further in Sec. 14.2. We can nevertheless be quite sure at this point that the physical mass of a portion of the universe will be *proportional* to its *volume*, due to the assumption of homogeneity.

Once we have found R as a function of ρ, or conversely, know $\rho = \rho(R)$, we may apply (13.4a) to find the time dependence of the radius R. These considerations show the freedom which we have in constructing cosmological models. We may prescribe the equation of state $p(\rho)$, the cosmological constant Λ, and the signature $k = \pm 1$ or 0. From this, all physical and geometric quantities are determined. However, we are not sure whether every model so obtained will be physically meaningful and appear in terms of real-valued functions.

It will be useful to summarize here the dimensionality of the different terms which appear in these equations: $R(t)$, which can be interpreted in a closed universe as a "radius," has the dimension of a length, Λ has the dimension of the inverse square of a length, and ρ is a mass density. Consequently, κ/c^2 has the dimension of a length divided by a mass, which is in agreement with its classical use. Theoretical physicists often like to reduce the number of dimensional constants used in a problem by arbitrarily setting some of them equal to 1 as long as compatibility permits. For instance, setting $\kappa/c^2 = 1$, we make mass and length units identical. Setting the speed of light $c = 1$ makes the time unit equal to the two others. All other constants of nature must then be expressible in terms of a single fundamental dimension; for instance, Planck's action constant has the dimension of the square of the fundamental dimension. We shall, however, remain explicit in our presentation and keep all constants in evidence, at the expense of a slight loss of brevity in the equations.

13.2 Static Models of the Universe

Formulas (13.4) and (13.5) give ρ and p as functions of $R(t)$. If we choose $R(t)$ to be a constant R independent of time, that is, a static universe, ρ and p will also be independent of time. Such an assumption does not lead to a red shift since it implies that $R_0 = R_e$ in (12.72). This makes any such model unsuitable for the description of the real physical universe unless one can explain the red shift from sources other than the space geometry. Moreover the blackbody radiation discussed in Sec. 12.1 finds no ready explanation in such a cosmological model. Let us, nevertheless, consider several particular cases of such models which are of historical interest.

Before entering into the mathematical development, we should explain why it is desirable to construct a model of a finite universe at all. Since the middle of the nineteenth century many physicists and astronomers had found disturbing paradoxes when they assumed an infinite and homogeneous universe. The simplest and most striking example is the famous Olbers paradox, which runs as follows: If we assume a matter distribution with constant average density throughout the universe and an equal behavior of matter at all points of the homogeneous universe, each volume element dV should emit radiation $\mu\, dV$ in all directions. Here μ is probably a very small constant, but it must be finite. It is also clear, according to classical notions, that the radiation density should decrease by an r^{-2} law, if r is the distance from the emitting volume element. Take, now, an observer at any given point of this universe. In a spherical shell of radius r and width dr around him lies a volume $4\pi r^2\, dr$, which radiates the total energy $\mu 4\pi r^2\, dr$. Of this, he receives only the amount proportional to $\mu(4\pi r^2\, dr)/r^2 = 4\pi\mu\, dr$. However, integrating over all values of r, we find that the total amount of energy radiated to our observer would be infinite. Thus the assumption of an infinite and uniform universe predicts an infinite brightness of the sky due to the summing of all contributions of the uniform matter distribution in the world. Similar paradoxes come from considerations of gravitational phenomena which are also mass-proportional and obey the inverse-square law. Hence, from the moment that Riemann indicated the possibility of a non-Euclidean geometry in a finite spherical world, the implications for cosmology were very seriously considered. In particular, since general relativity theory studies the close relations between geometry and matter, it was natural that, from the beginning, the question of finiteness of the universe should arise. We begin with the oldest and crudest models.

In the static case, Eqs. (13.4a) and (13.5a) become

$$\frac{8\pi\kappa}{c^2}\rho = -\Lambda + \frac{3k}{R^2} \tag{13.13a}$$

$$\frac{4\pi\kappa}{c^2}\left(\rho + \frac{3p}{c^2}\right) = \Lambda \tag{13.13b}$$

We can distinguish two cases, according to the value given to Λ.

Consider first $\Lambda = 0$; then, from (13.13*b*), the universe must be empty if we assume ρ and p to be nonnegative. Equation (13.13*a*) then implies that $k = 0$. That is, this case corresponds to a flat empty universe in which special relativity applies everywhere. On the other hand, to preserve the attractive possibility of a closed spherical universe ($k = 1$), one can also satisfy Eq. (13.13*b*) by assuming that the relation $\rho + 3p/c^2 = 0$ holds. This implies a large negative pressure $p = -\rho c^2/3$. The origin of such a negative pressure clearly cannot be easily explained as due to random residual velocities. It would therefore be necessary to postulate a new physical phenomenon with no observational justification to retain this model. Thus the model seems to be merely a mathematical construction of little physical significance. It leads, however, to a relation between R and ρ; from (13.13*a*)

$$R^2 = \frac{3kc^2}{8\pi\kappa\rho} \tag{13.14}$$

To keep the correct signature of the four-dimensional metric, R^2 must be positive, and therefore k must be $+1$. Thus one obtains a spherical universe of radius

$$R = c\left(\frac{3}{8\pi\kappa\rho}\right)^{1/2} \tag{13.15}$$

This yields $R \cong 10^{10}$ to 10^{11} light years for $\rho = 10^{-29}$ to 10^{-31} g/cm^3. In spite of its rather unphysical origin, this result for R is not physically unreasonable.

We now turn to the case $\Lambda \neq 0$, which gives more freedom in the parameters and allows a somewhat less artificial model. Although such a model probably cannot give a good description of the universe because of the absence of an intrinsic red shift, it is interesting to look at the order of magnitude that it predicts for Λ. Equations (13.4*a*) and (13.4*b*) give, for any positive pressure,

$$\frac{k}{R^2} < \Lambda < \frac{3k}{R^2} \tag{13.16}$$

Thus Λ has the order of magnitude of the inverse square of the "radius" of the universe. Therefore it cannot play an important role in the description of phenomena inside the solar system, as we indicated in Sec. 12.2. Equation (13.5*b*) gives in the static case

$$R^2 = \frac{kc^2}{4\pi\kappa(\rho + p/c^2)} \tag{13.17}$$

Therefore k must be equal to $+1$. Since, physically, ρ is much larger than p/c^2, R is roughly the same as for $\Lambda = 0$ [Eq. (13.14)].

Historically, it was to obtain such testable predictions linking physical density and the size of the universe that Einstein introduced the cosmological constant Λ. Only later came Hubble's discovery and the derivation of the general form of the Robertson-Walker metric. At the same time nonstatic solutions of Einstein's equations were found which did not necessitate the introduction of the cosmological constant Λ. This was a great relief to Einstein, who never liked to make use of this constant which had the "logical right" to enter the equations, but which Einstein considered arbitrary and aesthetically undesirable.

Before closing this section, we should point out another strong objection to the static universe as described by Eqs. (13.13). It was pointed out by Lemaître (1931) that the static solutions to these equations are unstable. That is, under certain perturbations of the density ρ, the system would enter an evolution away from the original static state. Since small perturbations are always present in the real universe, the existence of an unstable static universe appears very improbable.

13.3 Nonstatic Models of the Universe

In nonstatic models one considers a time-dependent $R(t)$, and therefore one can expect in this case to be able to obtain a description of the red shift. Logically, any functional form for $R(t)$ is possible if it leads to a red shift and allows ρ and p to be positive. Thus, in particular, $R(t)$ must be an increasing function, at least for the present time and observable past times. Indeed, in first approximation, observation gives the value of the logarithmic derivative of $R(t)$ at the present time through (12.83). In nonstatic models we still have the freedom of making Λ equal to zero or not.

In the case $\Lambda = 0$, Einstein's equations (13.4*a*) and (13.5*a*) become

$$\frac{8\pi\kappa}{c^2}\rho = \frac{3}{R^2c^2}R'^2 + \frac{3k}{R^2} \tag{13.18a}$$

$$\frac{4\pi\kappa}{c^2}\left(\rho + \frac{3p}{c^2}\right) = -\frac{3R''}{c^2R} \tag{13.18b}$$

The first equation yields a very interesting relation between the present density of the universe, Hubble's constant, and the parameter k. Rearrangement gives

$$\rho - \frac{3H^2}{8\pi\kappa} = \left(\frac{3c^2}{8\pi\kappa R^2}\right)k \tag{13.18c}$$

since $H = R_0'/R_0$ and $\rho = \rho_0$ in the present epoch. Thus a knowledge of accurate values for ρ_0 and H would determine the sign of k, and therefore the hyperspherical, Euclidean, or pseudo-hyperspherical character of the three-dimensional world. With the presently accepted value of $H^{-1} = (5.6 \pm 0.6) \times 10^{17}$ s, the left side is $\rho_0 - (2.1 \pm 0.5) \times 10^{-29}$ g/cm^3. As we discussed in Sec. 12.1, ρ_0 is at present not well known observationally, and we can only make the weak statement $\rho_0 \gtrsim 10^{-31}$ g/cm^3; thus as yet we cannot use (13.19) to determine k. Therefore we shall consider solutions for all three possibilities in this section.

Under the reasonable assumption that ρ and p are positive, the second equation shows that $R'' < 0$; that is, the expansion must be decelerated and $R(t)$ cannot have a minimum or an inflection point. Clearly the deceleration parameter q_0 must be positive, which is in agreement with present observations. To be able to obtain a definite expression for $R(t)$ as a function of time, we shall make the physically reasonable assumption that the pressure term can be neglected. This appears to be observationally justified in the present epoch. Then, by elimination of ρ between Eqs. (13.18a) and (13.18b), we obtain a single differential equation for $R(t)$.

$$2RR'' + R'^2 + kc^2 = 0 \tag{13.19}$$

This equation is equivalent to

$$(RR'^2)' + kc^2R' = 0 \tag{13.20}$$

which admits the first integral

$$R'^2 = \frac{D_0 - kR}{R}c^2 \tag{13.21}$$

Inserting (13.21) into (13.19), we obtain

$$R'' = -\frac{D_0}{2R^2}c^2 \tag{13.22}$$

Thus the constant D_0 has to be positive to give $R'' < 0$ as required by (13.18b). We can determine D_0 by again using (13.18b), which gives

$$D_0 = \left(\frac{4\pi}{3}R^3\rho\right)\frac{2\kappa}{c^2} \tag{13.23}$$

The term in parentheses is clearly proportional to the total mass of the universe from (12.70) for the case $k = 1$ and is equal to the mass of a sphere of size R for the Euclidean case, $k = 0$. Moreover it is conserved during the evolution of the pressureless universe, as already noted in (13.11). We therefore label it, for all values of k, by the symbol $\hat{M}$. Thus (13.23) may be written as

$$D_0 = \frac{2\hat{M}\kappa}{c^2} \tag{13.24}$$

which is formally the same as the expression giving the Schwarzschild radius (see Sec. 6.8).

We now consider the solutions of (13.21) in the three cases $k = +1$, 0, -1. Consider first $k = +1$. By setting $R = D_0 \sin^2 \tau(t)$, we can solve (13.21) parametrically in terms of τ to obtain

$$\begin{aligned} ct &= \frac{D_0}{2}[2\tau - \sin 2\tau] \\ R &= \frac{D_0}{2}[1 - \cos 2\tau] \end{aligned} \tag{13.25}$$

The radius R as a function of t describes a cycloid (see Fig. 13.1). It begins at zero at $t = \tau = 0$, corresponding to a "big bang," or explosive birth of the universe; of course that initial epoch cannot be quantitatively well described by our present considerations since we are neglecting pressure. The radius then increases, reaches a maximum value of $D_0 = 2\kappa\hat{M}/c^2$ at $t = D_0\pi/2c$, and then contracts again to $R = 0$. Note the curious feature that the universe is always inside a Schwarzschild radius, defined in the present case as D_0; this should not be construed to mean that the density is always high (see Exercise 13.4).

The theoretical quantities D_0 and R (at the present epoch) can readily be related to the observable quantities H and q_0 for this model. From (13.21) and (13.22) and the definition of q_0 (12.82') we obtain

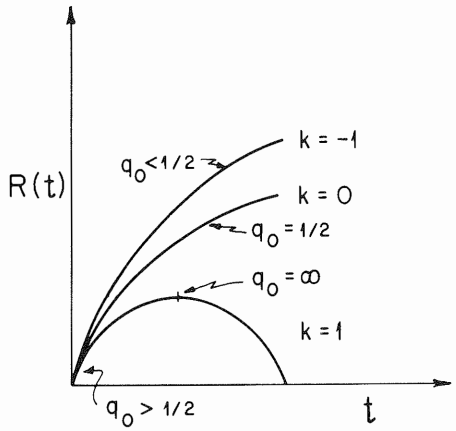


Fig. 13.1
Behavior of the dust-filled Robertson-Walker universe.

$$2q_0 = \frac{D_0}{D_0 - R} \qquad \frac{R}{D_0} = 1 - \frac{1}{2q_0} \tag{13.26}$$

Thus q_0 ranges from $\frac{1}{2}$ for $R = 0$ to infinity at $R = D_0$, consistent with the observational value 1.0 ± 0.5. From the identification of H in (12.86) we then also obtain

$$D_0 = \frac{2q_0}{(2q_0 - 1)^{3/2}} \frac{c}{H} \tag{13.27}$$

It is thus possible in principle to determine observationally all the properties of this model, which is indeed consistent with all present astronomical measurements (see Exercise 13.6).

Next let us consider the case $k = 0$. In this case (12.21) can be solved particularly easily, and we obtain

$$R^{3/2} - R_0^{3/2} = \tfrac{3}{2}\sqrt{D_0}\, c(t - t_0) \tag{13.28}$$

This solution allows a radius equal to zero at a time

$$t_1 = t_0 - \frac{2R_0^{3/2}}{3c\sqrt{D_0}}$$

By shifting the time origin to t_1 we obtain the simpler form

$$R = At^{2/3} \qquad A = (\tfrac{3}{2}c)^{2/3} D_0^{1/3} \tag{13.29}$$

This describes a universe in continuous expansion with an explosive birth at time $t = 0$, like the hyperspherical case (see Fig. 13.1). We can easily obtain the deceleration parameter from (13.21) and (13.22) as

$$q_0 = \frac{-R_0''}{R_0 R_0'^2} = \frac{1}{2} \tag{13.30}$$

We cannot obtain an expression for D_0 in terms of the observables H and q_0 but may instead obtain an expression for ρ_0 in terms of H from (12.18*c*)

$$\rho_0 = \frac{3H^2}{8\pi\kappa} = 2.1 \pm 0.5 \times 10^{-29} \text{ g/cm}^3 \tag{13.31}$$

Since these numbers are not inconsistent with the presently determined values of q_0 and ρ_0, the model is viable at present. This solution, with $\Lambda = k = 0$, is the original nonstatic solution proposed by Friedman (1922). It was the first solution leading to a physically acceptable cosmological solution of Einstein's equations without a cosmological constant. Since it is desirable in theoretical physics to describe nature with a minimum number of universal constants, there is justification for considering this solution to be preferable to preceding ones.

We come now to the last case, $k = -1$. Equation (13.21) becomes, in this case,

$$\frac{R'^2}{c^2} = \frac{D_0}{R} + 1 \tag{13.32}$$

To solve this, we substitute, in analogy with the case $k = 1$,

$$R = D_0 \sinh^2 \tau(t)$$

Then we obtain a solution very similar to (12.25):

$$\begin{aligned} ct &= \frac{D_0}{2}(\sinh 2\tau - 2\tau) \\ R &= \frac{D_0}{2}(\cosh 2\tau - 1) \end{aligned} \tag{13.33}$$

This shows that $R(t)$ increases monotonically from zero to infinity (see Fig. 13.1). Actually, this interesting fact could also have been read off from (13.32). Indeed, from (13.32) it is clear that R' remains of one sign and $R' > c$ if expansion starts. In fact, for large R, we see from (13.32) that the asymptotic behavior is

(12.34) $$R' = c \qquad R = ct$$

Therefore the model allows an expansion beyond D_0, the "Schwarzschild radius of the universe." The birth is explosive since $R' = \infty$ at $R = 0$, which is not quantitatively realistic due to the neglect of pressure. We can obtain interesting relations between D_0, R at the present epoch, q_0, and H, analogous to those obtained for the case $k = 1$. We leave it as an exercise to obtain the analogues of (13.26) and (13.27) and to show that $q_0 < \frac{1}{2}$. Such a value of q_0, and hence this model, is not inconsistent with the observed value of q_0.

In conclusion we may say that all the exploding models found in the case $\Lambda = 0$ are consistent with our present knowledge of the universe. An accurate observational value for q_0 could decide among the models since $q_0 > \frac{1}{2}$ for the hyperspherical model, $q_0 = \frac{1}{2}$ for the Euclidean model, and $q_0 < \frac{1}{2}$ for the pseudo-hyperspherical model; we can only say that $q_0 > \frac{1}{2}$ is somewhat favored by present data. The quantitative details of the explosive birth cannot be taken seriously due to the neglect of pressure. While it may be justifiable to neglect the pressure in the present state of the universe, this neglect is certainly not allowed for a highly condensed universe of small radius. However, if we kept the pressure term in the equations, we should have to specify the equation of state $p(\rho)$. It is difficult to say what this relation should be in a highly contracted universe, although there has been much speculation on the subject.

In the case $\Lambda \neq 0$ we have to deal with Eqs. (13.4) in all their generality. Originally, nonstatic solutions were introduced in order to allow one to dispense with the Λ-term in Einstein's equations. Therefore it is mainly for mathematical completeness that we mention here the possibility of nonstatic solutions, including the cosmological term. We shall present only one particular solution, which is of historical interest, having been the first nonstatic solution to be proposed.

This solution was given by de Sitter in 1917. It postulates that the rate of expansion of the universe is given exactly by the otherwise approximate relation

(13.35) $$\frac{R'}{R} = H$$

where H is Hubble's constant, as obtained in the limit of small astronomical distances. By elementary integration of (13.35), we obtain

(13.36) $$R = R_0 e^{Ht} \qquad R_0 = \text{const}$$

and the red shift versus parameter-distance relation (12.82) gives z proportional to l within $O(l^3)$. Such a model has no singularity. That is, $R(t)$ is neither zero nor infinite for finite times; it is unbounded in time and does not require a birth or a death. These properties give it a definite aesthetic appeal and mathematical convenience, but the consequences for ρ and p are drastic. Equations (13.5) reduce to

(13.37a) $$\frac{4\pi\kappa}{c^2}\left(\rho + \frac{3p}{c^2}\right) = \Lambda - \frac{3H^2}{c^2}$$

(13.37b) $$\frac{4\pi\kappa}{c^2}\left(\rho + \frac{p}{c^2}\right) = \frac{ke^{-2Ht}}{R_0^2}$$

The first equation indicates that $(\rho + 3p/c^2)$ is a constant, and the second indicates that $(\rho + p/c^2) \to 0$ for $t \to \infty$. Thus, with p and ρ positive, this model requires $\rho = p = 0$ at all times and also $k = 0$ from (13.37b). It therefore represents an empty universe with a time-dependent metric. It is Euclidean in space dimensions, but has an expanding scale. One can, however, still speak of a red shift if one thinks of a test atom and an observer situated in this otherwise empty universe; the wavelength of light emitted by the atom increases as the observer moves away from it. The actual universe might be considered as a set of local perturbations on a de Sitter geometry which is valid in the large. Locally, we could then solve the Schwarzschild problem and impose boundary conditions at infinity compatible with a de Sitter geometry.

The value of the cosmological constant is given by (13.37b):

(13.38) $$\Lambda = \frac{3H^2}{c^2} = \frac{1}{(1.8 \times 10^{10})^2}\ (\text{light years})^{-2}$$

It has the same order of magnitude as the value obtained in the static case, namely, the inverse square of the distance of the farthest objects seen today.

13.4 The Gödel Solution and Mach's Principle

In this section we disgress to discuss the Gödel solution, which does not have a Robertson-Walker type of metric. Our purpose is to present an interesting mathematical exercise and to illustrate in so doing that it is possible to find solutions of Einstein's equations without the simplifying mathematical assumptions used in Chap. 12; the exercise will also, we hope, clarify the relation between Mach's principle and general relativity theory.

Gödel (1949) showed that the following metric is compatible with an incoherent matter distribution:

$$ds^2 = (dx^0 + e^{\alpha x^1}\,dx^2)^2 - (dx^1)^2 - \tfrac{1}{2}e^{2\alpha x^1}(dx^2)^2 - (dx^3)^2 \tag{13.39}$$

The symbol α in this expression is a constant with the dimension of an inverse length. To verify that (13.39) is a solution we must compute the Christoffel symbols and the Einstein tensor corresponding to this metric. This is most easily done by obtaining the geodesic equations of (13.39). The Euler-Lagrange equations corresponding to the variational problem

$$\delta \int ds = 0 \tag{13.40}$$

are easily found to be

$$\begin{aligned}
&\frac{d}{ds}(\dot{x}^0 + e^{\alpha x^1}\dot{x}^2) = 0\\
-2&\frac{d}{ds}(\dot{x}^1) = 2\dot{x}^0\dot{x}^2\alpha e^{\alpha x^1} + (\dot{x}^2)^2\alpha e^{2\alpha x^1}\\
&\frac{d}{ds}\left[e^{\alpha x^1}\left(\dot{x}^0 + \frac{\dot{x}^2}{2}e^{\alpha x^1}\right)\right] = 0\\
&\frac{d}{ds}(\dot{x}^3) = 0
\end{aligned} \tag{13.41}$$

By slight rearrangement we can bring (13.41) into the simpler form

$$\begin{aligned}
&\ddot{x}^0 + 2\alpha\dot{x}^0\dot{x}^1 + \alpha e^{\alpha x^1}\dot{x}^1\dot{x}^2 = 0\\
&\ddot{x}^1 + \alpha e^{\alpha x^1}\dot{x}^0\dot{x}^2 + \frac{\alpha}{2}e^{2\alpha x^1}(\dot{x}^2)^2 = 0\\
&\ddot{x}^2 - 2\alpha e^{-\alpha x^1}\dot{x}^0\dot{x}^1 = 0\\
&\ddot{x}^3 = 0
\end{aligned} \tag{13.42}$$

From this we can find the nonvanishing Christoffel symbols in the same manner as in Chap. 6. There results

$$\begin{Bmatrix}0\\0\;\;1\end{Bmatrix} = \begin{Bmatrix}0\\1\;\;0\end{Bmatrix} = \alpha \qquad \begin{Bmatrix}0\\1\;\;2\end{Bmatrix} = \begin{Bmatrix}0\\2\;\;1\end{Bmatrix} = \frac{\alpha}{2}e^{\alpha x^1}$$

$$\begin{aligned}
&\begin{Bmatrix}1\\0\;\;2\end{Bmatrix} = \begin{Bmatrix}1\\2\;\;0\end{Bmatrix} = \frac{\alpha}{2}e^{\alpha x^1} \qquad \begin{Bmatrix}1\\2\;\;2\end{Bmatrix} = \frac{\alpha}{2}e^{2\alpha x^1}\\
&\begin{Bmatrix}2\\0\;\;1\end{Bmatrix} = \begin{Bmatrix}2\\1\;\;0\end{Bmatrix} = -\alpha e^{-\alpha x^1}
\end{aligned} \tag{13.43}$$

and all others are zero.

We now use the definition of $R_{\mu\nu}$:

$$R_{\mu\nu} = \begin{Bmatrix}\kappa\\\mu\;\;\kappa\end{Bmatrix}_{|\nu} - \begin{Bmatrix}\kappa\\\mu\;\;\nu\end{Bmatrix}_{|\kappa} + \begin{Bmatrix}\kappa\\\mu\;\;\lambda\end{Bmatrix}\begin{Bmatrix}\lambda\\\kappa\;\;\nu\end{Bmatrix} - \begin{Bmatrix}\lambda\\\kappa\;\;\lambda\end{Bmatrix}\begin{Bmatrix}\kappa\\\mu\;\;\nu\end{Bmatrix} \tag{13.44}$$

Each of the four terms on the right side of (13.44) is a 4×4 matrix with only a few nonzero components. Observe first that $\begin{Bmatrix}\lambda\\\kappa\;\;\lambda\end{Bmatrix} = \alpha$ for $\kappa = 1$ and zero for $\kappa \neq 1$. One then finds for the nonzero components the values

$$\begin{aligned}
&\begin{Bmatrix}\kappa\\0\;\;2\end{Bmatrix}_{|\kappa} = \begin{Bmatrix}\kappa\\2\;\;0\end{Bmatrix}_{|\kappa} = \frac{\alpha^2}{2}e^{\alpha x^1} \qquad &&\begin{Bmatrix}\kappa\\2\;\;2\end{Bmatrix}_{|\kappa} = \alpha^2 e^{2\alpha x^1}\\
&\begin{Bmatrix}\kappa\\0\;\;\lambda\end{Bmatrix}\begin{Bmatrix}\lambda\\0\;\;\kappa\end{Bmatrix} = -\alpha^2 &&\begin{Bmatrix}\kappa\\2\;\;\lambda\end{Bmatrix}\begin{Bmatrix}\lambda\\2\;\;\kappa\end{Bmatrix} = \frac{\alpha^2}{2}e^{2\alpha x^1}\\
&\begin{Bmatrix}\lambda\\\kappa\;\;\lambda\end{Bmatrix}\begin{Bmatrix}\kappa\\0\;\;2\end{Bmatrix} = \frac{\alpha^2}{2}e^{\alpha x^1} &&\begin{Bmatrix}\lambda\\\kappa\;\;\lambda\end{Bmatrix}\begin{Bmatrix}\kappa\\2\;\;2\end{Bmatrix} = \frac{\alpha^2}{2}e^{2\alpha x^1}
\end{aligned} \tag{13.45}$$

Substitution of these components into (13.44) immediately gives the contracted Riemann tensor

$$R_{\mu\nu} = -\alpha^2\begin{pmatrix}1 & 0 & e^{\alpha x^1} & 0\\ 0 & 0 & 0 & 0\\ e^{\alpha x^1} & 0 & e^{2\alpha x^1} & 0\\ 0 & 0 & 0 & 0\end{pmatrix} \tag{13.46}$$

The scalar $R^\mu{}_\mu$ is easily gotten by first inverting the metric tensor

$$g_{\mu\nu} = \begin{pmatrix}1 & 0 & e^{\alpha x^1} & 0\\ 0 & -1 & 0 & 0\\ e^{\alpha x^1} & 0 & \frac{1}{2}e^{2\alpha x^1} & 0\\ 0 & 0 & 0 & -1\end{pmatrix} \qquad \sqrt{-g} = \frac{e^{\alpha x^1}}{\sqrt{2}} \tag{13.47}$$

to get

$$g^{\mu\nu} = \begin{pmatrix} -1 & 0 & 2e^{-\alpha x^1} & 0 \\ 0 & -1 & 0 & 0 \\ 2e^{-\alpha x^1} & 0 & -2e^{-2\alpha x^1} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{13.48}$$

Raising one index of $R_{\mu\nu}$ and contracting then gives the simple result

$$R^{\mu}{}_{\mu} = -\alpha^2 \tag{13.49}$$

We have now calculated all the relevant geometric quantities connected with the metric (13.39).

Let us next compute the tensor $T_{\mu\nu}$. For an incoherent matter field at rest (we use co-moving coordinates), we know from Chap. 10 that

$$T^{\mu\nu} = \rho v^\mu v^\nu = \rho \delta^\mu{}_0 \delta^\nu{}_0 \tag{13.50}$$

Lowering indices with the metric tensor (13.47), we then have

$$T_{\mu\nu} = \rho g_{\mu 0} g_{\nu 0} = \rho \begin{pmatrix} 1 & 0 & e^{\alpha x^1} & 0 \\ 0 & 0 & 0 & 0 \\ e^{\alpha x^1} & 0 & e^{2\alpha x^1} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{13.51}$$

Note that this $T_{\mu\nu}$ is proportional to $R_{\mu\nu}$ in (13.46):

$$R_{\mu\nu} = -\frac{\alpha^2}{\rho} T_{\mu\nu} \tag{13.52}$$

Let us now collect the results of our calculations. The Einstein equations with the cosmological term may be written in the form

$$R_{\mu\nu} + (\Lambda - \tfrac{1}{2}R^\beta{}_\beta) g_{\mu\nu} = -\frac{8\pi\kappa}{c^2} T_{\mu\nu} \tag{13.53}$$

From (13.49) and (13.52) it is therefore evident that Gödel's metric is indeed a solution if

$$\Lambda = -\frac{\alpha^2}{2} \qquad \frac{\alpha^2}{\rho} = \frac{8\pi\kappa}{c^2} \tag{13.54}$$

Note that the cosmological constant Λ is necessary to have a nontrivial Gödel solution.

In summary we see that the Gödel metric (13.39) satisfies Einstein's equations if the conditions (13.54) are satisfied; moreover, it is evident from (13.54) that if α is equal to zero, both Λ and ρ must be zero, which implies that space is flat (as we showed in Chap. 8). Thus α can be considered as a parameter which measures the deviation of the space from flatness. This fact will be useful later in this section.

Let us note at this point an interesting feature of our result. The matter of the universe is at rest in the coordinate system we have used, so the system is a universal co-moving system. This leads us to a surprising conclusion: The energy-momentum tensor (13.50) is indeed precisely the same as was used in connection with the Einstein static universe in Sec. 13.2. Thus the field equations

$$R_{\mu\nu} + (\Lambda - \tfrac{1}{2}R) g_{\mu\nu} = C T_{\mu\nu} \qquad C = -\frac{8\pi\kappa}{c^2} \tag{13.55}$$

have two basically different solutions for the same $T_{\mu\nu}$. From the point of view of Mach's principle, we expect that the matter of the universe should uniquely determine the geometry of the universe. The situation noted above is clearly not consistent with this notion, which leads us to believe that Mach's principle is *not* built into general relativity via the field equations. Indeed, it appears that global notions (such as boundary conditions) must be added to general relativity to encompass Mach's principle. We shall say more about this later in the section.

We now wish to investigate some curious physical properties of the Gödel metric. First let us show that the world-lines which characterize matter at rest in Gödel's co-moving system (x^0,x^1,x^2,x^3) cannot be everywhere orthogonal to a one-parameter family of three-dimensional hypersurfaces. Note that this very important property of the metric intrinsically distinguishes the Gödel solution from any solution which admits a co-moving Gaussian coordinate system, that is, one for which $g_{00} = 1$ and $g_{0i} = 0$, such as the co-moving system of a Robertson-Walker metric. Indeed, if there existed such a family of hypersurfaces, one could mark off the s intervals of matter geodesics between hypersurfaces and thereby construct a co-moving Gaussian coordinate system by the method given in Sec. 2.4. It is thus evident that the above statement precludes the existence of a coordinate system in which there is a distinguished universal time-coordinate and in which the incoherent matter constituting the universe is at rest.

To demonstrate the statement of the preceding paragraph, suppose the contrary. That is, suppose that, imbedded in four-dimensional space, there is a family $\mathfrak{F}$ of three-dimensional hypersurfaces which are parametrized by λ and which have equations of the familiar form

$$F(x^\mu) - \lambda = 0 \qquad (F = \text{fixed function}) \tag{13.56}$$

If a vector dx^μ lies entirely within this surface, F will not change along it; that is, $dF = F_{|\mu}\, dx^\mu = 0$. Thus it is evident that one specific normal vector to the member of $\mathfrak{F}$ which contains the world-point x^μ is $F_{|\mu}(x^\mu)$. Thus any arbitrary vector field v_μ which is everywhere orthogonal to the members of $\mathfrak{F}$ may be written as

$$v_\mu = lF_{|\mu} \tag{13.57}$$

where l is an arbitrary scalar function. With this in mind, let us construct from an arbitrary vector v^μ the completely antisymmetric tensor

$$\begin{aligned} a_{\mu\nu\gamma} &\equiv \{v_\mu v_{\nu|\gamma}\} \\ &\equiv \frac{1}{3!}[v_\mu(v_{\nu|\gamma} - v_{\gamma|\nu}) + v_\nu(v_{\gamma|\mu} - v_{\mu|\gamma}) + v_\gamma(v_{\mu|\nu} - v_{\nu|\mu})] \end{aligned} \tag{13.58}$$

For the special case of the vector field v_μ given by (13.57), this tensor is easily seen by direct calculation to be identically zero. Thus a covariant necessary condition that a vector field v_μ be everywhere orthogonal to a one-parameter family $\mathfrak{F}$ of three-dimensional hypersurfaces is

$$a_{\mu\nu\gamma} = \{v_\mu v_{\nu|\gamma}\} = 0 \tag{13.59}$$

For the case of the Gödel solution, we are interested in the specific vector which represents matter at rest:

$$v^\mu = (1,0,0,0) \qquad v_\mu = (1,0,e^{\alpha x^1},0) \tag{13.60}$$

It is evident for this vector that

$$v_{\nu|\gamma} = \begin{cases} \alpha e^{\alpha x^1} & \text{for } \nu = 2,\ \gamma = 1 \\ 0 & \text{otherwise} \end{cases} \tag{13.61}$$

and therefore that the tensor $a_{\mu\nu\gamma}$ is

$$a_{\mu\nu\gamma} = \begin{cases} -\tfrac{1}{6}\alpha e^{\alpha x^1} & \text{even permutation of } 0,\ 1,\ 2 \\ \tfrac{1}{6}\alpha e^{\alpha x^1} & \text{odd permutation of } 0,\ 1,\ 2 \\ 0 & \text{otherwise} \end{cases} \tag{13.62}$$

Since $a_{\mu\nu\gamma}$ is not identically zero, we have completed our demonstration.

The tensor $a_{\mu\nu\gamma}$ which we have introduced above will also be useful in investigating the rotational behavior of a vector field. To see why the idea of rotation occurs in the first place in regard to the Gödel solution, let us recall a result of Chap. 4; we there obtained the metric corresponding to a flat space with a set of cylindrical coordinates r, φ, and z rotating about the z axis at a constant angular velocity w. The result (4.83) can be written as

$$ds^2 = \left(1 - \frac{w^2r^2}{c^2}\right)c^2\,dt^2 - dr^2 - r^2\,d\varphi^2 - dz^2 - 2wr^2\,dt\,d\varphi \tag{13.63}$$

We now compare this with the Gödel metric (13.39) written as

$$ds^2 = (dx^0)^2 - (dx^1)^2 - \tfrac{1}{2}e^{2\alpha x^1}(dx^2)^2 - (dx^3)^2 + 2e^{2\alpha x^1}\,dx^0\,dx^2 \tag{13.64}$$

There is an evident similarity between the forms of these two line elements. In particular, note that both have a cross term in the time interval. On the basis of this similarity, we shall make the tentative assumption that x^1 is a "radial" coordinate like r; x^2 is an "angular" coordinate like φ; and x^3 is an "axial" coordinate like z. (Actually Gödel carries out a transformation to a system r, φ, z, in which the correspondence between his line element and the "rotating flat space" line element (13.63) is made even more evident than above, but the computation is tedious and the result is not necessary for our purposes.) The formal similarity between (13.63) and (13.64) leads us to expect some sort of rotational behavior of the Gödel universe. Our problem is now to make such behavior explicit in a covariant way and justify the preceding intuitive notions.

To investigate the rotational character of the Gödel metric, let us use the tensor $a_{\mu\nu\gamma}$ introduced above to construct a new vector,

$$\Omega^\beta = c\,\frac{\epsilon^{\beta\mu\nu\gamma}}{\sqrt{-g}}\,a_{\mu\nu\gamma} = c\,\frac{\epsilon^{\beta\mu\nu\gamma}}{\sqrt{-g}}\,\{v_\mu v_{\nu|\gamma}\} \tag{13.65}$$

We thereby associate a vector Ω^β with a given vector v_μ. Because of the presence of the antisymmetric tensor $\epsilon^{\beta\mu\nu\gamma}/\sqrt{-g}$ in the definition, one might expect Ω^β to be closely related to the ordinary curl of v_μ. This is indeed the case, as one can easily see by working out the components of Ω^β in a flat space with the usual coordinates of special relativity so that $\sqrt{-g} = 1$. The contravariant vector $\Omega^\beta = (\Omega^0,\mathbf{\Omega})$ associated with $v_\mu = (v_0, -\mathbf{v})$ is then found to be

$$\Omega^0 = -c\mathbf{v}\cdot(\nabla \times \mathbf{v})$$
$$\boldsymbol{\Omega} = -cv_0(\nabla \times \mathbf{v}) - c(\mathbf{v}\times\dot{\mathbf{v}}) - c(\mathbf{v}\times\nabla)v_0 \tag{13.66}$$

To make the interpretation of Ω^β more transparent, consider the case of a vector field v_μ which represents a uniform field of counterclockwise rotation at angular velocity ω about the z axis of a three-dimensional Cartesian coordinate system x, y, z. That is, in these coordinates let

$$v_0 = 1 \qquad \mathbf{v} = v^i = \left(\frac{\omega}{c}\,y, -\frac{\omega}{c}\,x, 0\right) \tag{13.67}$$

By immediate calculation, using (13.66), we find that

$$\Omega^0 = 0$$
$$\boldsymbol{\Omega} = -c(\nabla \times \mathbf{v}) = (0,0,2\omega) \tag{13.68}$$

Thus the uniform rotation is characterized by

$$\Omega^\beta = (0,0,0,2\omega) \tag{13.69}$$

Now we apply the above interpretation of Ω^β to investigate the motion of co-moving matter in the Gödel universe. Note, first, that the covariant velocity vector v_μ of the co-moving matter field in a Robertson-Walker metric has the same components as the contravariant vector $v^\mu = (1,0,0,0)$, so both $a_{\mu\nu\gamma}$ and Ω^β are obviously zero for this field. However, for the Gödel metric, the matter field has an $a_{\mu\nu\gamma}$ given by (13.62), so one finds by direct calculation from (13.47) and (13.65) that

$$\Omega^\beta = (0,0,0,\sqrt{2}\,\alpha c) \tag{13.70}$$

This vector has precisely the same form as (13.69), which characterizes a uniform rotation about the z axis. We are thus led to interpret the result (13.70) as indicating that the co-moving matter in the Gödel universe possesses a constant *intrinsic* angular velocity $\omega = \alpha c/\sqrt{2}$ about the x^3 axis. This statement can also be expressed in terms of more basic parameters of the Gödel universe by the use of (13.54) and (13.49):

$$\omega = \frac{\alpha c}{\sqrt{2}} = 2\sqrt{\pi\kappa\rho} = c\left(\frac{-R^\mu{}_\mu}{2}\right)^{1/2} \tag{13.71}$$

The results of the preceding paragraph should leave the reader somewhat puzzled. We have shown formally that the co-moving matter of a Gödel universe undergoes an intrinsic uniform rotation. However, according to Mach's principle, the bulk matter of the universe should determine its geometry. Therefore, if the bulk matter of the universe is at rest in a particular coordinate system, one would expect that system to be inertial; for the Gödel universe this is not so. Furthermore, one must ask how the entire bulk matter of the universe can rotate; with respect to what does it rotate? To clarify these questions and put the problem in perspective, we shall study the motion of a test particle in a Gödel universe.

The equations of motion for a particle in a Gödel universe have already been obtained in (13.41). These equations can readily be partly integrated to give a new system

$$\dot{x}^0 + e^{\alpha x^1}\dot{x}^2 = a$$
$$e^{\alpha x^1}\left(\dot{x}^0 + \frac{\dot{x}^2}{2}\,e^{\alpha x^1}\right) = e^{\alpha x^1}\left(a - \frac{\dot{x}^2}{2}\,e^{\alpha x^1}\right) = b \tag{13.72}$$
$$\ddot{x}^1 + \alpha b\dot{x}^2 = 0 \qquad \dot{x}^1 + \alpha b x^2 = d$$
$$\dot{x}^3 = l$$

where a, b, d, and l are constants of integration. It is easy to see from these equations that a particle of matter at rest,

$$x^\mu = (s,A,B,C) \qquad \dot{x}^\mu = (1,0,0,0) \qquad A, B, C \text{ const} \tag{13.73}$$

follows a geodesic for the following choice of constants:

$$a = 1 \qquad l = 0 \qquad b = e^{\alpha A} \qquad d = \alpha b B \tag{13.74}$$

Let us ask what happens when a particle is not allowed to remain at rest but is given an initial velocity in the co-moving system. Specifically, suppose that we shoot a particle from the origin $x^i = 0$ in a radial direction toward some fixed target galaxy at the position $x^i = (A,0,0)$ (Fig. 13.2). The initial conditions are then

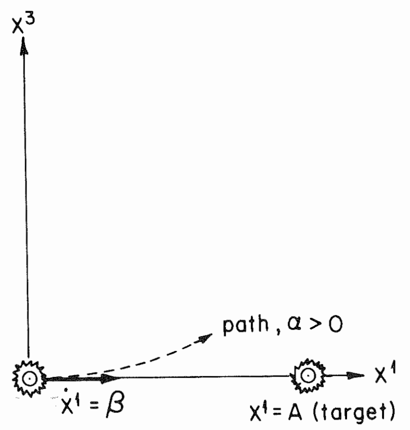


Fig. 13.2

$$x^0 = 0 \qquad \dot{x}^0 = 1 \qquad x^1 = 0 \qquad \dot{x}^1 = \beta$$
$$x^2 = 0 \qquad \dot{x}^2 = 0 \qquad x^3 = 0 \qquad \dot{x}^3 = 0 \tag{13.75}$$

which correspond to the constants

$$a = b = 1 \qquad d = \beta \qquad l = 0 \tag{13.76}$$

as one easily sees from Eqs. (13.72). Intuition and Mach's principle would then lead one to expect that, since matter determines geometry, the particle should travel in a straight line toward the distant target nebula. That is, the solution of (13.72) should have the form

$$x^\mu = (x^0(s), x^1(s), 0, 0)$$

We shall find, however, that this is not the case for the Gödel universe.

It is not difficult to solve the equations of motion (13.72) exactly for the initial conditions (13.75). However, the exact solution has a complicated form and is not very enlightening. We can learn all we need to know about the motion by supposing that α is a small parameter and then using perturbation theory. Recall that α is a measure of the nonflatness of space, so a perturbation approach is clearly well motivated. Let us begin by setting $\alpha = 0$ in (13.72). The solution of the resultant equations is elementary, and we obtain

$$x^\mu = (s,\beta s,0,0) \tag{13.77}$$

Thus, for flat space $\alpha = 0$, our conjecture in the preceding paragraph

is verified: The particle travels directly toward the distant target galaxy at constant coordinate velocity β.

Now suppose that α is small but nonzero, and expand equations (13.72) to first order in α:

$$\begin{aligned} \dot{x}^0 + \dot{x}^2 - 1 &= -\alpha x^1 \dot{x}^2 \\ \ddot{x}^2 &= 2\alpha x^1(1 - \dot{x}^2) \\ \dot{x}^1 - \beta &= -\alpha x^2 \\ \dot{x}^3 &= 0 \end{aligned} \tag{13.78}$$

Into these equations we substitute the solution (13.77) plus a first-order correction αy^μ:

$$\begin{aligned} x^0 &= s + \alpha y^0 \\ x^1 &= \beta s + \alpha y^1 \\ x^2 &= \alpha y^2 \\ x^3 &= \alpha y^3 \end{aligned} \tag{13.79}$$

Using only first-order terms in α, this gives a set of equations for the y^μ:

$$\begin{aligned} \dot{y}^0 + \dot{y}^2 &= 0 \\ \ddot{y}^2 &= 2\beta s \\ \dot{y}^1 &= 0 \\ \dot{y}^3 &= 0 \end{aligned} \tag{13.80}$$

Since the zeroth-order terms of (13.79) satisfy the initial conditions (13.75), all the y^μ and $\dot{y}^\mu$ are zero at $s = 0$. We thus arrive at the first-order solution

$$x^\mu = (s - \alpha\beta s^2, \beta s, \alpha\beta s^2, 0) \tag{13.81}$$

It is evident from (13.80) that the particle deviates from the ray

$$x^2 = x^3 = 0$$

and spirals outward instead of traveling directly toward the distant target galaxy, as it did for the case of flat space.

We are in a position to answer a few questions and come to some tentative and speculative conclusions concerning the Gödel universe.

First, we see that the matter of this universe does indeed rotate relative to "something"; the "something" is the path that a test particle follows if it is given an initial radial velocity. One may refer to the tangent of such a path as the "compass of inertia." Thus the compass of inertia rotates relative to the matter of the Gödel universe, or vice versa. Second, the interpretation of Mach's principle which we have mentioned above can be stated in more precise form: *The bulk matter* (or "fixed stars") *of the universe determines the compass of inertia, and the two cannot rotate relative to each other.* The Gödel solution, therefore, is consistent with general relativity theory, but not with this statement of Mach's principle; that is, general relativity does not imply Mach's principle as it is stated above. Thus it appears that ultimately Mach's principle must be incorporated into general relativity theory by the addition of some appropriate boundary or global condition.

13.5 The Steady-State Model of the Universe

In 1948 Bondi and Gold proposed a cosmological model which does not rely on Einstein's equations (Bondi and Gold, 1948). In their theory they do not postulate a specific link between matter and geometry in the form of a mathematical relation between $T^{\mu\nu}$ and $R^{\mu\nu}$, but instead postulate a constant energy density ρ and demand that an observer at any point in space-time find the same Hubble recession (12.2). All points in space-time thus have to be equivalent; this postulate is often called the "perfect cosmological postulate."

Put together, the above two requirements preclude the possibility of an energy conservation law. In fact, to keep the matter density ρ constant and to compensate for the expansion of the universe, it is necessary to assume that matter is created at a constant rate throughout space. To show this let us take an observer and draw a fixed sphere of radius r around him, using Euclidean geometry as an approximation. Because of the expansion of the scale of space, matter will flow out of this sphere at a rate $4\pi r^2 v\rho$, where v is the radial velocity of matter at the distance r. To compensate for this outflow, matter must be created inside the sphere at a constant rate $(4\pi/3)r^3Q$, where Q is the average rate of creation of matter density. Thus, in a steady-state universe, one must have

$$4\pi r^2 \rho v = \frac{4\pi}{3} r^3 Q \tag{13.82}$$

Using Hubble's law, with $T = H^{-1}$,

$$v = \frac{r}{T} \tag{13.83}$$

we obtain

$$Q = \frac{3\rho}{T} \tag{13.84}$$

which, translated into equivalent numbers of hydrogen atoms, gives $Q \simeq 10^{-(15\pm2)}$ hydrogen atoms/cm^3 year. Considering that ρ, so far as we know it, is within two orders of magnitude of 6×10^{-6} hydrogen atoms/cm^3, such a rate of matter creation certainly would be undetectable.

The main appeal of this model is that it has no singularity in its geometry; there is no beginning and no end of the universe. The constancy of T makes all times equivalent. However, the conceptual difficulty of a birth at the origin of time encountered in several earlier models is now replaced by the assumption of a continuous creation of matter, which may be equally hard to accept. To take this hypothesis seriously, we must also ask in what form this matter can be created. At this point it should be recalled that recent studies of Burbidge et al. (1957) have shown that the synthesis of elements in stars through successive nuclear reactions starting with free protons is theoretically possible; this theory explains the relative abundances of the elements in the universe rather well. In order that the steady-state theory remain consistent with these results, we therefore have to suppose that the matter is created in the form of hydrogen atoms, a somewhat awkward and artificial assumption.

Another serious objection to the steady-state theory is that it provides no ready explanation of the blackbody radiation discussed in Sec. 12.1. As a result the theory is now considered by most cosmologists as less viable than the evolving models discussed in Sec. 13.3. We include it here to illustrate how readily the mathematical structure of relativity theory can be altered to describe novel physical effects.

We shall now attempt to put the steady-state theory into a mathematical framework by the use of equations which are analogous to Einstein's equations without the cosmological term. It is impossible a priori to reconcile the steady-state theory with Einstein's equations as they now stand because of the energy conservation implied in these equations,

$$-\frac{8\pi\kappa}{c^2} T^{\mu\nu} = G^{\mu\nu} \tag{13.85}$$

since the right-hand side has a zero divergence. In the steady-state theory, the energy-momentum tensor cannot have a zero divergence because of the continuous creation of matter. However, one can write down equations of the same form as Einstein's equations by subtracting from $T^{\mu\nu}$ a tensor corresponding to the matter creation. This approach was proposed by Hoyle (1948), who wrote down modified Einstein equations of the form

$$-\frac{8\pi\kappa}{c^2}(T^{\mu\nu} - C^{\mu\nu}) = G^{\mu\nu} \tag{13.86}$$

Since $T^{\mu\nu}$ and $G^{\mu\nu}$ are symmetric, the tensor $C^{\mu\nu}$, which is thus introduced, must also be symmetric.

Equations (13.86) are not sufficient to determine completely the mathematical problem, so a subsidiary condition on the tensor $C^{\mu\nu}$ must be added. This can be done in the following more or less natural way. In the distinguished Gaussian coordinate system which we shall use, there is a naturally distinguished vector, $D^\mu = (1,0,0,0)$. Since the tensor $C^{\mu\nu}$ which is introduced in (13.86) is also a distinguished tensor in this Gaussian system, one might desire to relate D^μ and $C^{\mu\nu}$ to each other in order to keep the number of distinguished tensors to a minimum and thereby achieve a mathematical economy in the theory. The simplest way to relate a vector D^μ and a symmetric tensor $C^{\mu\nu}$ is via

$$C^{\mu\nu} = \tfrac{1}{2}A(D^{\mu\|\nu} + D^{\nu\|\mu}) \equiv \tfrac{1}{2}A(g^{\mu\alpha}D^\nu{}_{\|\alpha} + g^{\nu\alpha}D^\mu{}_{\|\alpha}) \tag{13.87}$$

where A is a constant scalar which we can later choose for our convenience. We thus add to the modified Einstein equations (13.86) the subsidiary requirement (13.87), with $D^\mu = (1,0,0,0)$. We shall see in the following paragraphs that the distinguished tensor thus defined does indeed put the steady-state theory into consistent mathematical form.

The postulates of homogeneity and equivalence of all world-points completely determine the metric. Indeed, the metric clearly must be of the Robertson-Walker type, and furthermore, because of the assumed constant rate of expansion of the universe, it must also be of the de Sitter type, with

$$R(t) = R_0 e^{t/T} \tag{13.88}$$

We shall, moreover, assume a priori that $k = 0$; that is, the three-dimensional space is Euclidean. Recall that the de Sitter universe can exist only for a vanishing energy density. In the present case, however, it is possible to obtain a model of the universe with a de Sitter type of metric and a nonvanishing constant matter density ρ, because of the presence of the matter-creation term in Eqs. (13.86). We shall write the Robertson-Walker metric in the form

$$ds^2 = dx_0^2 - \frac{R^2(t)}{R_0^2}(dx_1^2 + dx_2^2 + dx_3^2) \qquad x_0 = ct \qquad R(t) = R_0 e^{t/T} \tag{13.89}$$

Thus the metric tensor $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$ are given by

$$g_{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -\left(\frac{R}{R_0}\right)^2 & & \\ & & -\left(\frac{R}{R_0}\right)^2 & \\ & & & -\left(\frac{R}{R_0}\right)^2 \end{pmatrix}$$

$$g^{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -\left(\frac{R_0}{R}\right)^2 & & \\ & & -\left(\frac{R_0}{R}\right)^2 & \\ & & & -\left(\frac{R_0}{R}\right)^2 \end{pmatrix} \tag{13.90}$$

Let us now attack the mathematical problem of finding the most general tensor $C^{\mu\nu}$ such that the modified Einstein equations (13.86) hold for a universe with a Robertson-Walker metric of type (13.89) and in which there is a constant and uniform density ρ of matter which is anchored to co-moving coordinates. We shall later use the subsidiary condition (13.87) to single out the simplest and most appropriate choice of $C^{\mu\nu}$. We have, actually, a very easy task, since the tensor $G^{\mu\nu}$ is completely determined by the metric (13.89), and the energy-momentum tensor for incoherent matter anchored to co-moving coordinates is also uniquely given by $T^{\mu\nu} = \rho\delta^\mu{}_0\delta^\nu{}_0$. Thus the modified Einstein equations readily determine $C^{\mu\nu}$.

Let us compute $G^\mu{}_\nu$ first. Using $G^0{}_0$ and $G^1{}_1 = G^2{}_2 = G^3{}_3$ as conveniently displayed in (13.4) with the specific choices of $k = 0$, $R = R_0 e^{t/T}$, we arrive at

$$G^\mu{}_\nu = -\frac{3}{c^2T^2}\delta^\mu{}_\nu \tag{13.91}$$

Then, by using $T^\mu{}_\nu = \rho\delta^\mu{}_0\delta^0{}_\nu$, we can solve the modified Einstein equations for $C^\mu{}_\nu$:

$$(13.92)\quad C^\mu{}_\nu = T^\mu{}_\nu + \frac{c^2}{8\pi\kappa} G^\mu{}_\nu$$

$$= \begin{pmatrix} \rho - \dfrac{3}{8\pi\kappa T^2} & & & \\ & -\dfrac{3}{8\pi\kappa T^2} & & \\ & & -\dfrac{3}{8\pi\kappa T^2} & \\ & & & -\dfrac{3}{8\pi\kappa T^2} \end{pmatrix}$$

Raising one index with $g^{\mu\nu}$ given in (13.90), we also have $C^{\mu\nu}$:

$$(13.93)\quad C^{\mu\nu} = \begin{pmatrix} \rho - \dfrac{3}{8\pi\kappa T^2} & & & \\ & \left(\dfrac{R_0}{R}\right)^2 \dfrac{3}{8\pi\kappa T^2} & & \\ & & \left(\dfrac{R_0}{R}\right)^2 \dfrac{3}{8\pi\kappa T^2} & \\ & & & \left(\dfrac{R_0}{R}\right)^2 \dfrac{3}{8\pi\kappa T^2} \end{pmatrix}$$

We have not yet utilized the subsidiary condition (13.87). In order to apply it we first need to calculate the nonvanishing Christoffel symbols and then compute the tensor $D^\mu{}_{\|\nu}$. Using the Robertson-Walker metric (13.89), we find the nonzero Christoffel symbols as usual by obtaining the differential equations of the geodesic lines. The result is

$$(13.94)\quad \begin{aligned} \begin{Bmatrix} 0 \\ i\ j \end{Bmatrix} &= \frac{R'R}{cR_0^2}\delta_{ij} = \left(\frac{R}{R_0}\right)^2 \frac{1}{cT}\delta_{ij} \\ \begin{Bmatrix} i \\ 0\ j \end{Bmatrix} &= \frac{R'}{Rc}\delta_{ij} = \frac{1}{cT}\delta_{ij} \end{aligned}$$

The tensor $D^\mu{}_{\|\nu}$ is therefore

$$(13.95)\quad D^\mu{}_{\|\nu} = D^\mu{}_{|\nu} + \begin{Bmatrix} \mu \\ \alpha\ \nu \end{Bmatrix} D^\alpha$$

$$= \begin{pmatrix} 0 & & & \\ & \dfrac{1}{cT} & & \\ & & \dfrac{1}{cT} & \\ & & & \dfrac{1}{cT} \end{pmatrix}$$

Since both $D^\mu{}_{\|\nu}$ and $D^{\mu\|\nu}$ are clearly symmetric, we see from (13.87) that the subsidiary condition is $C^\mu{}_\nu = AD^\mu{}_{\|\nu}$; by comparison of (13.93) and (13.95), we thus obtain

$$(13.96)\quad \rho = \frac{3}{8\pi\kappa T^2} \qquad A = \frac{-3c}{8\pi\kappa T}$$

so that Hubble's constant is the only universal constant which still appears in the theory.

Lastly, let us investigate the divergence of $T^{\mu\nu} = \rho\delta^\mu{}_0\delta^\nu{}_0$, which in the present theory must be equal to the divergence of $C^{\mu\nu}$, since $G^{\mu\nu}$ is divergenceless. An easy calculation using the Christoffel symbols (13.94) gives

$$(13.97)\quad T^{\mu\nu}{}_{\|\nu} = T^{\mu\nu}{}_{|\nu} + \begin{Bmatrix} \mu \\ \alpha\ \nu \end{Bmatrix} T^{\nu\alpha} + \begin{Bmatrix} \nu \\ \alpha\ \nu \end{Bmatrix} T^{\alpha\mu} = \frac{3\rho}{cT}(1,0,0,0)$$

and

$$(13.98)\quad C^{\mu\nu}{}_{\|\nu} = C^{\mu\nu}{}_{|\nu} + \begin{Bmatrix} \mu \\ \alpha\ \nu \end{Bmatrix} C^{\nu\alpha} + \begin{Bmatrix} \nu \\ \alpha\ \nu \end{Bmatrix} C^{\mu\alpha} = \frac{3\rho}{cT}(1,0,0,0)$$

The two divergence vectors are therefore equal as the theory requires. There is another interesting feature of the result (13.98). One can interpret $T^{\mu\nu}{}_{\|\nu} = C^{\mu\nu}{}_{\|\nu}$ as the source vector of the energy-momentum tensor $T^{\mu\nu}$ in the same way as $s^\mu = F^{\mu\nu}{}_{\|\nu}$ is the source vector of the Minkowski tensor $F^{\mu\nu}$, as indicated in (4.58). The explicit form (13.98) for this vector therefore indicates that matter density is created at rest in the co-moving coordinate system with a rate $Q = 3\rho/T$. This is exactly the same relation as (13.84), which we found previously by heuristic arguments. Moreover, we can now relate the creation rate Q to Hubble's constant T by using the relation between ρ and T in (13.96):

$$(13.99)\quad Q = \frac{3\rho}{T} = \frac{9}{8\pi\kappa T^3}$$

In conclusion we can say that we have succeeded in setting up a consistent mathematical scheme for a steady-state cosmological theory with a Robertson-Walker type of metric and continuous creation of matter at a rate $Q = 3\rho/T$. The relation (13.96) between the density ρ and Hubble's constant is the same as that found by Friedman, according to (13.31), and is compatible with the present observational data as we indicated in Sec. 13.3.

13.6 Converse of the Apparent Magnitude–Red Shift Problem

In this section we shall consider the converse mathematical problem of Sec. 12.6: If one were given an accurate empirical relation between the apparent magnitude m_0 and the red shift z, what conclusions could be drawn concerning the geometry compatible with the relation? In particular, could one determine the function $R(t)$ of a corresponding Robertson-Walker metric directly from observational data?

As an illustrative example we select the case of the apparent magnitude–red shift relation (12.103), which we suppose now to be exact. This is equivalent to supposing Hubble's law to be exact, as we discussed in Sec. 12.6. From (12.97) and (12.103) we obtain by comparison

$$\frac{R_0}{R_e}\frac{u}{1+(k/4)u^2} = Bz \tag{13.100}$$

where B is a properly chosen constant. In view of (12.98) we can also assert that

$$\frac{u}{1+(k/4)u^2} = B\frac{z}{1+z} = \frac{B}{R_0}(R_0 - R_e) \tag{13.101}$$

On the other hand, we have a fundamental relation between R and u from (12.72), (12.73), and (12.106),

$$\begin{aligned}\int_{t_e}^{t_0}\frac{c\,dt}{R(t)} &= \int_0^{u_0}\frac{du}{1+(k/4)u^2}\\ &= \frac{2}{\sqrt{k}}\arctan\left(\tfrac{1}{2}\sqrt{k}\,u_0\right) = \frac{2}{\sqrt{k}}\theta_0\end{aligned} \tag{13.102}$$

where the auxiliary variable $\theta = \arctan(\frac{1}{2}\sqrt{k}\,u)$ has been introduced for convenience. With this definition of θ we have

$$\sin 2\theta = \frac{2\tan\theta}{1+\tan^2\theta} = \frac{\sqrt{k}\,u}{1+(k/4)u^2} \tag{13.103}$$

Hence, (13.102) and (13.103) lead to the integral relation

$$\frac{u}{1+(k/4)u^2} = \frac{1}{\sqrt{k}}\sin\left\{\sqrt{k}\int_{t_e}^{t_0}\frac{c\,dt}{R(t)}\right\} \tag{13.104}$$

and (13.101) becomes a nonlinear integral equation:

$$R(t_0) - R(t_e) = \frac{R(t_0)}{B\sqrt{k}}\sin\left\{\sqrt{k}\int_{t_e}^{t_0}\frac{c\,dt}{R(t)}\right\} \tag{13.105}$$

This condition determines the value of $R(t)$ if we assume the precise relation (12.103) between apparent magnitude and red shift.

We have still the freedom to choose k and to prescribe the type of geometry desired. The simplest choice is evidently $k = 0$, which we shall pursue a little further. From the series development of $\sin\alpha$ it follows that (13.105) reduces in this case to

$$R(t_0) - R(t_e) = \frac{R(t_0)}{B}\int_{t_e}^{t_0}\frac{c\,dt}{R(t)} \tag{13.106}$$

which is a linear integral equation of much simpler form. It can even be reduced to an ordinary differential equation. We hold t_0 fixed and differentiate (13.106) on both sides with respect to the variable emission time t_e. We find, then, that

$$\frac{cR(t_0)}{B} = R(t_e)R'(t_e) = \frac{d}{dt_e}\left[\tfrac{1}{2}R(t_e)^2\right] \tag{13.107}$$

This has the obvious solution

$$R(t_e) = R(t_0)\left[1 - 2\frac{(t_0 - t_e)c}{BR(t_0)}\right]^{1/2} \tag{13.108}$$

where the arbitrary constant of integration has been chosen to fit the initial condition

$$R(t_e) = R(t_0) \qquad \text{for } t_e = t_0 \tag{13.109}$$

If we set $t_e = t_0$ in (13.107), we find that $c/B = R'(t_0)$, and using the approximate relation (12.86), we may connect the constant B with the Hubble constant H by the equation

$$BR(t_0) = \frac{c}{H} \tag{13.110}$$

Observe, however, that when $k = 0$, we have, by (13.101) and (12.72),

$$\frac{z}{1+z} = \frac{HR_0}{c} u = \frac{HR_0}{c} l \tag{13.111}$$

From this formula we can calculate the precise relation between the markers L as defined by Hubble's law $z = HL/c$ and l defined from the Robertson-Walker metric, instead of the first approximation $L = R_0 l$, which is frequently used in astronomy.

The Robertson-Walker geometry with $R(t)$ given by

$$R(t) = R(t_0)[1 + 2(t - t_0)H]^{1/2} \tag{13.112}$$

has several interesting features. We see that, at the moment $t_i < t_0$, determined by the condition

$$t_i = t_0 - \frac{1}{2H} \tag{13.113}$$

the radius of the universe is zero. If we measure time from this starting point, i.e., shift our time origin so that $t_i = 0$, we find $t_0 = 1/2H$ and the $R(t)$ law reduces to the simpler form

$$R(t) = R\left(\frac{1}{2H}\right)(2Ht)^{1/2} \tag{13.114}$$

This law implies an interesting pressure-density relation. Indeed, the Einstein field equations for $\Lambda = k = 0$ and for a Robertson-Walker metric imply, by (13.4) and (13.114),

$$\frac{8\pi\kappa}{c^2}\rho = \frac{3}{4c^2t^2} \qquad \frac{8\pi\kappa}{c^2}\frac{p}{c^2} = \frac{1}{4c^2t^2} \tag{13.115}$$

and thus the relation between pressure and density is

$$\rho = 3\left(\frac{p}{c^2}\right) = \frac{3}{32\pi\kappa t_e^2} \tag{13.116}$$

Recall, however, that the relation

$$\frac{p}{\rho} = \frac{c^2}{3} \tag{13.117}$$

characterizes electromagnetic radiation. Thus the "molecules" which comprise the "gas" of this universe would move at the speed of light and must therefore be photons, neutrinos, or other massless particles. It is generally supposed that p may be neglected compared to ρ in the present epoch, as we did in Sec. 13.3, but we really know very little about the density of such things as the massless particles in the universe, as mentioned in Sec. 12.1. We therefore cannot rule out the pressure-density relation (13.117). Moreover by comparing (12.103), on which this model is based, and (12.111) we see that $q_0 = 1$. This is consistent with the observational value of $q_0 = 1.0 \pm 0.5$. Thus this model is consistent with present data.

From the discussion in Sec. 13.3 and the above results we see that the present observational data are not yet adequate to allow us to reach definite conclusions about the validity of various models.

Exercises

13.1 In (13.9) we showed that the energy balance under cosmic evolution is preserved. Prove as stated in the text that entropy is also constant during the evolution.

13.2 Set up a system of natural units in which κ and c are both numerically equal to 1. One may then equate length and mass units. What mass corresponds to a length of 1 cm? What is the geometric mass $\kappa M/c^2$ of a proton, the earth, the sun, and the universe?

13.3 (*continued*) Discuss how one may work always in natural units without converting back to conventional units at the end of a theoretical derivation.

13.4 What is the approximate density of a system equal in size to its Schwarzschild radius if that radius is 1 cm (approximately the Schwarzschild radius of the earth), 3 km (approximately the Schwarzschild radius of the sun), and 10^{11} light years (approximately the Schwarzschild radius of the universe)?

13.5 A cycloid is the curve described by a point on the periphery of a rolling wheel; using this picture, give a geometrical interpretation of the parameter 2τ in (13.25) as an angle.

13.6 Use the value $q_0 = 1.0 \pm 0.5$ given in the text to obtain a numerical value for D_0 in the hyperspherical universe and a numerical value for

R in the present epoch. What are the observational limits on these quantities? How long will the universe last before $R = 0$ again? What is its present density?

13.7 Why can we not obtain D_0 and R in terms of H and q_0 for the Euclidean universe, $k = 0$?

13.8 Obtain analogues of (13.26) and (13.27) for the case $k = -1$, and also show that $q_0 < \frac{1}{2}$.

13.9 At what distance from us, in the nonstatic solutions, do receding galaxies reach the speed of light as observed at the present epoch? What is the observational effect, and is such a velocity of recession in consistent with any of the fundamental ideas of relativity?

Problems

13.1 What explanation of the red shift in Hubble's law could be given for a static nonevolving universe?

13.2 What explanation of the blackbody radiation could be given in a static nonevolving universe?

13.3 Show that the static universes defined by Eqs. (13.13) are unstable.

13.4 Discuss the physical density of a Robertson-Walker universe as opposed to the coordinate density ρ.

13.5 What is the Petrov type of the Gödel solution? (See Exercises 10.8 and 12.6).

Bibliography

Bondi, H. (1960): "Cosmology," 2d ed., Cambridge.

Bondi, H., and T. Gold (1948): The Steady State Theory of the Expanding Universe, *Monthly Notices Roy. Astron. Soc.*, **108**:252–270.

Burbidge, E. M., G. R. Burbidge, W. A. Fowler, and F. Hoyle (1957): Synthesis of Elements, *Rev. Mod. Phys.*, **29**:547–650.

Friedman, A. (1922): Über die Krümmung des Raumes, *Z. Physik*, **10**:377–386.

Gödel, K. (1949): An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation, *Rev. Mod. Phys.*, **21**:447–450.

Gödel, K. (1950): Rotating Universes in General Relativity Theory, *Proc. Intern. Congr. Math. Cambridge, Mass.*, **1**:175–181.

Heckmann, O., and E. Schücking (1959): Newtonsche und Einsteinsche Kosmologie: Andere kosmologische Theorien, in Encyclopedia of Physics, vol. 53, Berlin-Göttingen-Heidelberg, pp. 489–537.

Heckman, O., and E. Schücking (1962): Relativistic Cosmology, in L. Witten (ed.), "Gravitation: An Introduction to Current Research," New York, pp. 438–469.

Hoyle, F. (1948): A New Model for the Expanding Universe, *Monthly Notices Roy. Astron. Soc.*, **108**:372–382.

Hoyle, F. (1949): On the Cosmological Problem, *Monthly Notices Roy. Astron. Soc.*, **109**:365–371.

Hoyle, F., and A. Sandage (1956): The Second Order Term in the Redshift Magnitude Relation, *Publ. Astron. Soc. Pacific*, **68**:301–307.

Humason, M. L., N. U. Mayall, and A. R. Sandage (1956): Redshifts and Magnitudes of Extragalactic Nebulae, *Astron. J.*, **61**:97–162.

Lemaître, G. (1931): The Expanding Universe, *Monthly Notices Roy. Astron. Soc.*, **91**:490–501.

McVittie, G. C. (1956): "General Relativity and Cosmology," London.

Minkowski, R. (1960): A New Distant Cluster of Galaxies, *Publ. Astron. Soc. Pacific*, **72**:354–356.

Robertson, H. P. (1955): The Theoretical Aspects of the Nebular Redshift, *Publ. Astron. Soc. Pacific*, **67**:82.

Sandage, A. (1961): The Ability of the 200-inch Telescope to Discriminate between Selected World Models, *Astrophys. J.*, **133**:355–392.

Schrödinger, E. (1956): "Expanding Universe," Cambridge, England.

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.

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 (14.14a) 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 (14.14b) 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 (14.14b) and utilize (14.14c) 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:

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

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

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

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

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

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 (14.25*d*). To obtain the remaining metric function ν it is necessary to solve (14.25*e*); 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 (14.25*b*) immediately to obtain

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

and therefore, by (14.25*d*),

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

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

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

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

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

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 (14.25*e*) or (14.21) between pressure, density, and the metric function $\nu(r)$. We obtain

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

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

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

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

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

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

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

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

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

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

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

the expression for ρ obtained from (14.46)

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

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

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

The square of this relation yields

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

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

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

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

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

which, inserted in (14.48), gives

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

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

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

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

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

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.

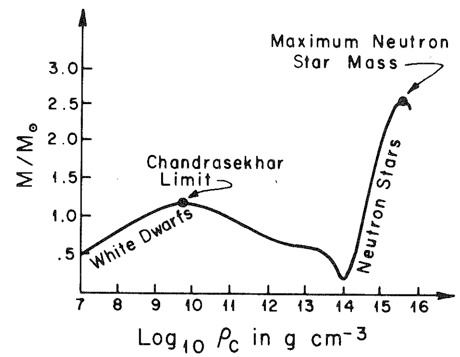
With an equation of state and a given central density ρ_c, Eqs. (14.25*b*) and (14.25*c*) allow one to construct an interesting function, the mass of the star in terms of its central density. For this purpose it is necessary to integrate these equations for the functions p, ρ, and m, which will depend 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 (14.70*a*); it is also clear from (14.70*b*) 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, $\rho =$ 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

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

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

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

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 (15.2*c*) independent of the function $E(r)$. $E(r)$ must be determined along with ν and λ from (15.2*a*) and (15.2*b*). 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 (15.2*b*) can now be solved for E. We substitute (15.8) into (15.2*b*) 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 (15.2*a*), 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

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

We can now proceed to solve (15.2*a*), 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 (15.14*a*) by $e^{-(\nu-\lambda)}$ and add to the second (15.14*b*) 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

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

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

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

Observe that differentiating this gives rise to (15.14a), so (15.14a), (15.14b), 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

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

which integrates immediately to give

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

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

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

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

$$E = \frac{\epsilon}{r^2} \tag{15.22}$$

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 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).

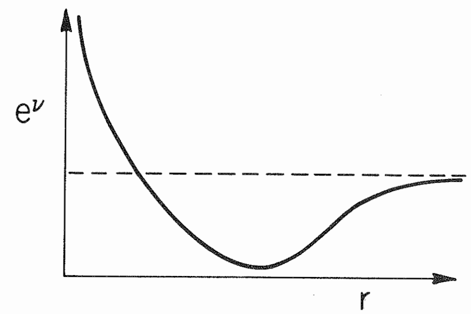


Fig. 15.1

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

$$\Gamma^\alpha{}_{\beta\gamma} = -\left\{ {\alpha \atop \beta\ \gamma} \right\} + g^{\sigma\alpha}[g_{\sigma\beta}\varphi_\gamma + g_{\sigma\gamma}\varphi_\beta - g_{\beta\gamma}\varphi_\sigma] \tag{15.30}$$

Here $\left\{ {\alpha \atop \beta\ \gamma} \right\}$ 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

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

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

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

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

$$\varphi_\alpha \equiv 0 \qquad \Gamma^\alpha{}_{\beta\gamma} = -\begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix} \tag{15.33}$$

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

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

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 $-\begin{Bmatrix} \alpha \\ \beta\ \gamma \end{Bmatrix}$ 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 $\varphi_\lambda = \frac{1}{2}(\log f)_{|\lambda}$:

$$\hat{\Gamma}^\alpha{}_{\beta\gamma} = \Gamma^\alpha{}_{\beta\gamma} \tag{15.35}$$

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

$$\hat{\varphi}_\lambda = \frac{1}{2}\frac{\partial}{dx^\lambda}\log f \tag{15.36}$$

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$S \equiv I + tF + t^2F^2 + \cdots \tag{15.100}$$

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

$$(I - tF)S = S(I - tF) = I \tag{15.101}$$

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

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

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

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

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:

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

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

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

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

We define next the characteristic polynomial of the matrix F as

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

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

$$|I - tF| = t^n \left| \frac{1}{t} I - F \right| = t^n \phi\left(\frac{1}{t}\right) = 1 + a_{n-1}t + \cdots + a_0 t^n \tag{15.107}$$

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

$$M = F^n + a_{n-1}F^{n-1} + \cdots + a_0 I \tag{15.108}$$

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:

$$\phi(F) = F^n + a_{n-1}F^{n-1} + \cdots + a_0 I = 0 \tag{15.109}$$

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

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

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

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 τ, τ, $-\tau$, and $-\tau$ is similar to a diagonal matrix $\mathfrak{M}$,*

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

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

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

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

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

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):

$$\mathrm{Tr}\,\Gamma = 0 \tag{15.161}$$

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

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

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:

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

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,

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

Thus, in the sense of the foregoing paragraph, we can refer to $\partial x^\mu/\partial \bar{x}^0$ as v^μ and to $(\partial x^\nu/\partial \bar{x}^0)R_{\mu\nu}$ as w_μ. Then (15.176) can be written as

$$-\bar{R}_{00} = v^\mu w_\mu \tag{15.178}$$

where v^μ and w^μ are timelike, spacelike, or null together. It is clear from (15.178) that the covariant requirement that $R_{\alpha\beta}$ carry v_α into a vector w_α such that $v^\alpha w_\alpha \geq 0$ is equivalent to the requirement that $\bar{R}_{00}$ be negative-definite. This can be interpreted physically by saying that $-R_{\alpha\beta}$ must carry the *forward* light cone into itself and the *backward* light cone into itself. The negativeness of the component R_{00} is thereby guaranteed in a covariant way for any real coordinate system:

$$R_{00} \leq 0 \qquad \text{(real coordinates)} \tag{15.179}$$

One must be careful to include the above restriction to a real coordinate system. For instance, if one goes from real x^0 to imaginary $\bar{x}^0$, then $\frac{\partial x^0}{\partial \bar{x}^0}\frac{\partial x^0}{\partial \bar{x}^0}$ is clearly negative, and the statement cannot be true.

The conditions (15.163), (15.165), and (15.179) were first obtained by Rainich in 1925, using somewhat different matrix methods than we have used. They are usually referred to as the *algebraic Rainich conditions.*

The Rainich conditions are purely algebraic, in the sense that they follow entirely from the form of the matrix $T_{\mu\nu}$ as constructed from the antisymmetric matrix $F_{\mu\nu}$ at a single world-point P. They do not involve the change in any quantities as one moves from world-point to world-point. We next need to take into account the fact that $F_{\mu\nu}$ obeys the Maxwell equations, which we now write in the form

$$F^{\mu\nu}{}_{\|\nu} = 0 \qquad {}^*F^{\mu\nu}{}_{\|\nu} = 0 \tag{15.180}$$

At P the coordinates are locally geodesic, and therefore the Christoffel symbols vanish. This allows us to drop the distinction between covariant and ordinary derivatives of first order. Furthermore, since the metric tensor is the Kronecker $\delta_{\mu\nu}$, it is possible to ignore index position. Maxwell's equations at P then can be written as

$$F_{\mu\nu|\nu} = 0 \qquad {}^*F_{\mu\nu|\nu} = 0 \tag{15.181}$$

where we have used the Einstein summation convention without regard for index position.

Our aim now is to study how Eqs. (15.181) reflect themselves in properties of the $T_{\mu\nu}$ and $R_{\mu\nu}$ tensors. Observe that, given the energy-momentum tensor $T_{\mu\nu}$, we can obtain the generating tensor $F_{\mu\nu}$ only up to a parameter α according to (15.156). At every point x^ν the value of $\lambda_1^2 - \lambda_2^2$ is determined by the energy-momentum tensor $T_{\mu\nu}$, while α may be chosen as a function of x^ν. Algebraically, this α field could be completely incoherent at different world-points. However, since $F_{\mu\nu}$ must satisfy Eq. (15.181), the α becomes a determined point function for which a differential system in terms of $T_{\mu\nu}$ or $R_{\mu\nu}$ can be given. The integrability condition on this system leads to a set of differential conditions on the tensor R, which was first discovered by Misner and Wheeler in 1957.

In order to derive the Misner-Wheeler equations we observe that, if the electromagnetic tensor is written

$$F_{\mu\nu} = (r \cosh \alpha)p_{\mu\nu} + (r \sinh \alpha)q_{\mu\nu} \tag{15.182}$$

as in (15.156), where $r = \sqrt{\lambda_1^2 - \lambda_2^2}$, then its dual is

$${}^*F_{\mu\nu} = (r \cosh \alpha)q_{\mu\nu} + (r \sinh \alpha)p_{\mu\nu} \tag{15.183}$$

This follows from considering the effect of the * operation on the simple forms

$$\bar{p}_{\mu\nu} = \begin{pmatrix} iJ & 0 \\ 0 & 0 \end{pmatrix} \qquad \bar{q}_{\mu\nu} = \begin{pmatrix} 0 & 0 \\ 0 & iJ \end{pmatrix} \tag{15.184}$$

where the interchange of $\bar{p}$ and $\bar{q}$ under the * operation is evident. But this operation is coordinate-invariant and must hold also in any system where p and q are gotten from $\bar{p}$ and $\bar{q}$ by an orthogonal transformation Q. The general result (15.183) is thereby assured.

The choice of our special coordinate system is arbitrary up to an orthogonal transformation. This allows us to assume without loss of generality that at P the matrices p and q are precisely $\bar{p}$ and $\bar{q}$ in (15.184); that is, $Q = I$. It follows, then, from (15.143) that the matrix

$$\Gamma_{\mu\nu} = c^2 T_{\mu\nu}$$

at P is

$$\bar{\Gamma}_{\mu\nu} = \tfrac{1}{2}r^2 \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \qquad r^2 = \lambda_1^2 - \lambda_2^2 \tag{15.185}$$

The Maxwell equations (15.181) are, then, from (15.182) and (15.183),

$$(15.186a)\quad F_{\mu\nu|\nu} = \tilde{p}_{\mu\nu}(r\cosh\alpha)_{|\nu} + \tilde{q}_{\mu\nu}(r\sinh\alpha)_{|\nu} + (r\cosh\alpha)p_{\mu\nu|\nu} + (r\sinh\alpha)q_{\mu\nu|\nu} = 0$$

$$(15.186b)\quad {}^*F_{\mu\nu|\nu} = \tilde{q}_{\mu\nu}(r\cosh\alpha)_{|\nu} + \tilde{p}_{\mu\nu}(r\sinh\alpha)_{|\nu} + (r\cosh\alpha)q_{\mu\nu|\nu} + (r\sinh\alpha)p_{\mu\nu|\nu} = 0$$

(Clearly, if we differentiate the $p_{\mu\nu}$ and $q_{\mu\nu}$, we must drop the tilde.) To simplify this form of Maxwell's equations, we define at P local vectors Π_μ and K_μ and scalars A and B by

$$(15.187)\quad \Pi_\mu = p_{\mu\nu|\nu} \qquad K_\mu = q_{\mu\nu|\nu} \qquad A = r\cosh\alpha \qquad B = r\sinh\alpha$$

Then (15.186) reads

$$(15.188a)\quad \tilde{p}_{\mu\nu}A_{|\nu} + \tilde{q}_{\mu\nu}B_{|\nu} + A\Pi_\mu + BK_\mu = 0$$

$$(15.188b)\quad \tilde{q}_{\mu\nu}A_{|\nu} + \tilde{p}_{\mu\nu}B_{|\nu} + AK_\mu + B\Pi_\mu = 0$$

Now we multiply (15.188*a*) by the matrix $\tilde{p}$ and (15.188*b*) by $\tilde{q}$, add the two, and use (15.152) to obtain

$$(15.189)\quad A_{|\mu} + \tilde{p}_{\mu\nu}\Pi_\nu A + \tilde{p}_{\mu\nu}K_\nu B + \tilde{q}_{\mu\nu}K_\nu A + \tilde{q}_{\mu\nu}\Pi_\nu B = 0$$

Whence, by definition of A and B,

$$(15.190)\quad A[\tilde{p}_{\mu\nu}\Pi_\nu + \tilde{q}_{\mu\nu}K_\nu + (\log r)_{|\mu}] + B[\tilde{q}_{\mu\nu}\Pi_\nu + \tilde{p}_{\mu\nu}K_\nu + \alpha_{|\mu}] = 0$$

Similarly, multiplying (15.188*a*) by $\tilde{q}$ and (15.188*b*) by $\tilde{p}$ and adding, we obtain

$$(15.191)\quad A[\tilde{q}_{\mu\nu}\Pi_\nu + \tilde{p}_{\mu\nu}K_\nu + \alpha_{|\mu}] + B[\tilde{p}_{\mu\nu}\Pi_\nu + \tilde{q}_{\mu\nu}K_\nu + (\log r)_{|\mu}] = 0$$

Equations (15.190) and (15.191) have the form $Ax + By = 0$ and $Bx + Ay = 0$. These equations have a nonzero solution for x and y only if the determinant of the coefficients, $A^2 - B^2$, is zero. Since, however, $A^2 - B^2 = \lambda_1^2 - \lambda_2^2 \neq 0$, the only solution is $x = y = 0$, so the coefficients of A and B in (15.190) and (15.191) are both zero. Thus we obtain the following differential conditions on α and r:

$$(15.192a)\quad \alpha_{|\mu} = -(\tilde{q}_{\mu\nu}\Pi_\nu + \tilde{p}_{\mu\nu}K_\nu)$$

$$(15.192b)\quad (\log r)_{|\mu} = -(\tilde{p}_{\mu\nu}\Pi_\nu + \tilde{q}_{\mu\nu}K_\nu)$$

The next step in our development is to calculate the vectors Π_μ and K_μ and to use the results to put (15.192*a*) into more interesting form. In order to do this we first have to find the derivatives of the matrices p and q. At the world-point P, the matrices p and q are precisely $\tilde{p}$ and $\tilde{q}$. If we consider, however, the world-point $P(\epsilon,\mu)$, obtained by moving a small amount ϵ in the direction of the μth coordinate axis, we shall have matrices $p^{(\mu)}$ and $q^{(\mu)}$, which differ from $\tilde{p}$ and $\tilde{q}$ by a small rotation in space-time corresponding to an orthogonal matrix $Q_{(\mu)}$. Let us represent $Q_{(\mu)}$ by a series

$$(15.193)\quad Q_{(\mu)} = I + \epsilon C_{(\mu)} + \epsilon^2 D_{(\mu)} + \cdots$$

and use the orthogonality condition on $Q_{(\mu)}$, $Q^T_{(\mu)}Q_{(\mu)} = I$, to find that

$$(15.194)\quad I = I + \epsilon(C_{(\mu)} + C^T_{(\mu)}) + \cdots$$

Equating powers of ϵ we get

$$(15.195)\quad C^T_{(\mu)} = -C_{(\mu)}$$

which is simply the well-known fact that infinitesimal rotations are generated by antisymmetric matrices. We therefore have

$$(15.196)\quad Q^T_{(\mu)} = Q^{-1}_{(\mu)} = I - \epsilon C_{(\mu)} + \cdots$$

The above result will enable us to calculate the derivative of the matrix p, $p_{|\mu}$, at P. Using (15.193) and (15.196), we obtain $p^{(\mu)}$ at $P(\epsilon,\mu)$ as

$$(15.197)\quad p^{(\mu)} = Q^T_{(\mu)}\tilde{p}Q_{(\mu)} = \tilde{p} + \epsilon(\tilde{p}C_{(\mu)} - C_{(\mu)}\tilde{p}) + O(\epsilon^2)$$

The derivative of $p^{(\mu)}$ at P is thus easily obtained by writing

$$(15.198)\quad \frac{p^{(\mu)} - \tilde{p}}{\epsilon} = (\tilde{p}C_{(\mu)} - C_{(\mu)}\tilde{p}) + O(\epsilon)$$

and taking the limit as $\epsilon \to 0$,

$$(15.199)\quad p_{|\mu} = \lim_{\epsilon\to 0}\frac{p^{(\mu)} - \tilde{p}}{\epsilon} = \tilde{p}C_{(\mu)} - C_{(\mu)}\tilde{p}$$

Let us write C_μ in terms of 2×2 matrices as

$$C_{(\mu)} = \begin{pmatrix} a_{(\mu)} & b_{(\mu)} \\ c_{(\mu)} & d_{(\mu)} \end{pmatrix} \tag{15.200}$$

where the antisymmetry of $C_{(\mu)}$ implies that $a^T_{(\mu)} = -a_{(\mu)}$, $d^T_{(\mu)} = -d_{(\mu)}$, and $b^T_{(\mu)} = -c_{(\mu)}$. Then (15.199) takes the simple form

$$p_{|\mu} = i\begin{pmatrix} 0 & Jb_{(\mu)} \\ -c_{(\mu)}J & 0 \end{pmatrix} \tag{15.201}$$

In precisely the same manner we obtain the derivative of q at P:

$$q_{|\mu} = i\begin{pmatrix} 0 & -b_{(\mu)}J \\ Jc_{(\mu)} & 0 \end{pmatrix} \tag{15.202}$$

Later we shall need the derivative of $\Gamma_{\mu\nu} = c^2 T_{\mu\nu}$, so we shall obtain it now while it is most convenient; recall that Γ in terms of p and q is given by

$$\Gamma = \tfrac{1}{2}r^2(p^2 - q^2) \tag{15.203}$$

Using (15.194), (15.195), (15.201), and (15.203), we find that at P

$$\begin{aligned} \Gamma_{|\mu} &= 2\tilde{\Gamma}(\log r)_{|\mu} + r^2(\tilde{p}p_{|\mu} - \tilde{q}q_{|\mu}) \\ &= (\log r)_{|\mu}r^2\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} + r^2\begin{pmatrix} 0 & b_{(\mu)} \\ -c_{(\mu)} & 0 \end{pmatrix} \end{aligned} \tag{15.204}$$

In the results (15.201) to (15.204), only the 2×2 matrices $b_{(\mu)}$ and $c_{(\mu)} = -b^T_{(\mu)}$ occur. Let us write these as

$$b_{(\mu)} = \begin{pmatrix} k_\mu & l_\mu \\ m_\mu & n_\mu \end{pmatrix} \qquad c_{(\mu)} = -\begin{pmatrix} k_\mu & m_\mu \\ l_\mu & n_\mu \end{pmatrix} \tag{15.205}$$

Substituting these into (15.201) to (15.204), we obtain

$$p_{|\mu} = i\begin{pmatrix} 0 & Jb_{(\mu)} \\ -c_{(\mu)}J & 0 \end{pmatrix} = i\left(\begin{array}{cc|cc} 0 & 0 & m_\mu & n_\mu \\ 0 & 0 & -k_\mu & -l_\mu \\ \hline -m_\mu & k_\mu & 0 & 0 \\ -n_\mu & l_\mu & 0 & 0 \end{array}\right) \tag{15.206a}$$

$$q_{|\mu} = i\begin{pmatrix} 0 & -b_{(\mu)}J \\ Jc_{(\mu)} & 0 \end{pmatrix} = i\left(\begin{array}{cc|cc} 0 & 0 & l_\mu & -k_\mu \\ 0 & 0 & n_\mu & -m_\mu \\ \hline -l_\mu & -n_\mu & 0 & 0 \\ k_\mu & m_\mu & 0 & 0 \end{array}\right) \tag{15.206b}$$

$$\Gamma_{|\mu} = (\log r)_{|\mu}r^2\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} + r^2\left(\begin{array}{cc|cc} 0 & 0 & k_\mu & l_\mu \\ 0 & 0 & m_\mu & n_\mu \\ \hline k_\mu & m_\mu & 0 & 0 \\ l_\mu & n_\mu & 0 & 0 \end{array}\right) \tag{15.206c}$$

Thus we finally arrive at

$$\Pi_\mu = p_{\mu\nu|\nu} = i\begin{pmatrix} m_3 + n_4 \\ -k_3 - l_4 \\ -m_1 + k_2 \\ -n_1 + l_2 \end{pmatrix} \tag{15.207a}$$

$$K_\mu = q_{\mu\nu|\nu} = i\begin{pmatrix} l_3 - k_4 \\ n_3 - m_4 \\ -l_1 - n_2 \\ k_1 + m_2 \end{pmatrix} \tag{15.207b}$$

We thus have achieved our goal of obtaining an explicit form for the vectors K_μ and Π_μ in terms of the elements of the generating matrix C_μ. Substitute now these terms (15.207) into (15.192a) to obtain

$$\alpha_{|\mu} = \begin{pmatrix} n_3 - m_4 \\ k_4 - l_3 \\ l_2 - n_1 \\ m_1 - k_2 \end{pmatrix} \tag{15.208}$$

One should note an extraordinary thing at this point: The vector $\alpha_{|\mu}$ is composed of the same components k_μ, l_μ, m_μ, and n_μ which occur in $\Gamma_{|\mu}$. This is a very important observation, and the only problem remaining before we obtain the final Wheeler-Misner condition is to express this correspondence in a covariant manner. With this goal in mind, let us compute the covariant vector

$$v_\lambda = \frac{\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}\Gamma^{\beta\mu|\nu}\Gamma_\mu{}^\gamma}{\Gamma_{\rho\kappa}\Gamma^{\rho\kappa}} \tag{15.209}$$

in our special coordinate system at the world-point P. [Recall that, in Eq. (3.25*a*), we showed that $\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}$ is a tensor.] Computation of the denominator is immediate from (15.172):

$$\tilde{\Gamma}_{\rho\kappa}\tilde{\Gamma}^{\rho\kappa} = \mathrm{Tr}\,(\Gamma^2) = r^4 \tag{15.210}$$

Next observe that the numerator in (15.209) takes the simple form in our special coordinate system

$$\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}\Gamma^{\beta\mu\|\nu}\Gamma_\mu{}^\gamma = \epsilon_{\lambda\nu\beta\gamma}\Gamma_{\beta\mu|\nu}\tilde{\Gamma}_{\mu\gamma} \tag{15.211}$$

which, from (15.206*c*), is

$$\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}\Gamma^{\beta\mu\|\nu}\Gamma_\mu{}^\gamma = \epsilon_{\lambda\nu\beta\gamma}\left[2\tilde{\Gamma}_{\beta\mu}(\log r)_{|\nu} + r^2\begin{pmatrix} 0 & b_{(\mu)} \\ -c_{(\mu)} & 0 \end{pmatrix}\right]\tilde{\Gamma}_{\mu\gamma} \tag{15.212}$$

Note, however, that the Rainich condition (15.164) tells us that $\tilde{\Gamma}_{\beta\mu}\tilde{\Gamma}_{\mu\gamma}$ is a multiple of the identity matrix; thus, since $\epsilon_{\lambda\nu\beta\gamma}$ is antisymmetric in β and γ, the first term of (15.212) obviously vanishes and we are left with

$$\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}\Gamma^{\beta\mu\|\nu}\Gamma_\mu{}^\gamma = \epsilon_{\lambda\nu\beta\gamma}\begin{pmatrix} 0 & b_{(\mu)} \\ -c_{(\mu)} & 0 \end{pmatrix}\tilde{\Gamma}_{\mu\gamma}r^2 \tag{15.213}$$

Substituting now $\tilde{\Gamma}_{\mu\nu}$ from (15.185) and $b_{(\mu)}$ and $c_{(\mu)}$ from (15.205) into this expression, we obtain

$$\Gamma^{\beta\mu\|\nu}\Gamma_\mu{}^\gamma = \tfrac{1}{2}r^4\begin{pmatrix} 0 & 0 & -k_\nu & -l_\nu \\ 0 & 0 & -m_\nu & -n_\nu \\ k_\nu & m_\nu & 0 & 0 \\ l_\nu & n_\nu & 0 & 0 \end{pmatrix} \tag{15.214}$$

and hence, by a slightly tedious but elementary calculation,

$$v_\lambda = \frac{\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}\Gamma^{\beta\mu\|\nu}\Gamma_\mu{}^\gamma}{\Gamma_{\rho\kappa}\Gamma^{\rho\kappa}} = \begin{pmatrix} n_3 - m_4 \\ k_4 - l_3 \\ l_2 - n_1 \\ m_1 - k_2 \end{pmatrix} \tag{15.215}$$

Comparing this with (15.208), we see that we have achieved our goal, for at P in our special coordinate system,

$$\alpha_{|\lambda} = v_\lambda \equiv \frac{\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}\Gamma^{\beta\mu\|\nu}\Gamma_\mu{}^\gamma}{\Gamma_{\rho\kappa}\Gamma^{\rho\kappa}} \tag{15.216}$$

This, however, is written in tensor form and is thus true in general. In view of the simple relation (15.160) between the tensors $\Gamma^{\mu\nu}$ and $T^{\mu\nu}$, we may replace in this identity $\mathbf{\Gamma}$ by $\mathbf{T}$. By use of the Einstein equations (15.2*a*), this may be put into a purely geometrical form:

$$\alpha_{|\lambda} = v_\lambda = \frac{\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}R^{\beta\mu\|\nu}R_\mu{}^\gamma}{R_{\rho\kappa}R^{\rho\kappa}} \tag{15.217}$$

which is our final result.

The basic equation of Wheeler and Misner follows immediately from (15.217), for the condition that (15.217) be integrable is

$$v_{\lambda|\tau} - v_{\tau|\lambda} = 0 \tag{15.218}$$

This constitutes six additional differential conditions on $R_{\mu\nu}$ in order that it correspond to an electromagnetic field.

In conclusion, let us restate the conditions of Rainich, Wheeler, and Misner from (15.163), (15.165), (15.179), and (15.218):

$$R_{\mu\nu}R^\nu{}_\alpha = \tfrac{1}{4}(R_{\tau\beta}R^{\tau\beta})g_{\mu\alpha} \tag{15.219a}$$

$$R_{00} \leq 0 \qquad \text{(in real coordinates)} \tag{15.219b}$$

$$R^\mu{}_\mu = 0 \tag{15.219c}$$

$$v_{\lambda|\tau} - v_{\tau|\lambda} = 0 \qquad v_\lambda = \frac{\sqrt{-g}\,\epsilon_{\lambda\nu\beta\gamma}R^{\beta\mu\|\nu}R_\mu{}^\gamma}{R_{\rho\kappa}R^{\rho\kappa}} \tag{15.219d}$$

These four sets of conditions are the basis of Wheeler's "already unified" field theory. Note that, since (15.219*a*) and (15.219*d*) are quadratic in $R_{\mu\nu}$, only the inequality (15.219*b*) serves to determine the overall sign of $R_{\mu\nu}$.

Let us now take the following new point of view: Suppose we are given the system (15.219) to begin with and are asked to calculate $g_{\mu\nu}$. In order to simplify the problem we might introduce a fictitious new tensor $F_{\mu\nu}$ which satisfies Maxwell's equations and then define a symmetric tensor $T_{\mu\nu}$ as in (15.2*a*). Finally, we should set $R_{\mu\nu}$ proportional to $T_{\mu\nu}$

and try to solve the resulting system. All this is possible by virtue of (15.219). [Such a procedure would reproduce system (15.2) by retracing our steps in this section which led to (15.219).] This point of view makes it clear that one might wish to think of $F_{\mu\nu}$ as only a convenient mathematical construct.

The procedure described in the preceding paragraph is reminiscent of a standard artifice in two-dimensional potential theory. If we wish to solve Laplace's equation for the unknown $u(x,y)$,

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{15.220}$$

it is often very useful to introduce an auxiliary function $v(x,y)$ by the definition

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{15.221}$$

That such a function exists follows from the differential equation (15.220) itself. It also follows that v satisfies Laplace's equation, $\nabla^2 v = 0$. Thus, instead of dealing with one second-order differential equation for u, we have achieved a reduction and splitting into the two first-order differential equations (15.221) at the price of introducing the fictitious unknown function $v(x,y)$. In almost all applications of Laplace's equation the function $v(x,y)$ has a simple and important physical interpretation. For example, in fluid dynamics, where u is the velocity potential of a flow, v will play the role of the stream function.

We may interpret the electromagnetic field in the already unified field theory in a way which is analogous to that of $v(x,y)$ in potential theory. Consider (15.219) as the basic system of equations for the $R_{\mu\nu}$ tensor. This entirely geometrical system is, unfortunately, very nonlinear and difficult to handle mathematically. In order to solve it, introduce the $F_{\mu\nu}$ tensor as previously indicated. This splits the nonlinear system (15.219) into a different set of equations (15.2) which are linear in $R_{\mu\nu}$ and not so difficult to solve. In Wheeler's viewpoint the simplification achieved by this reduction is so tremendous that for more than a century physicists have ascribed a physical reality to $F_{\mu\nu}$ and have assumed the existence of an electromagnetic field independent of the metric structure of space-time.

There remains now one further problem in this approach to a unified field theory: How do we describe and explain charged matter, i.e., the sinks and sources of the field? The basic relations in (15.219) are valid only at places where the field is regular and no charge exists. Thus it

would appear necessary to admit singularities in the field as in Sec. 15.1. There is, however, one possibility of geometrizing away even the singularities. As mentioned above, the relations (15.219) lead back to Maxwell's equations, and these in turn lead to the concept of lines of force which can end only at singularities of the field. We need these singularities, therefore, to serve as sinks and sources for lines of force; more mathematically, we need singularities only in order to circumvent the uniqueness theorems on the solutions of partial-differential-equation systems, and thereby exclude the trivial solution of an empty and static world. Wheeler accomplishes this end without introducing singularities by endowing the world with an appropriate novel topology. Suppose, for example, that in first approximation the world is a sphere in four-space. One may deform the sphere by adding a handle between world-points P_1 and P_2 as shown in Fig. 15.2. Such a deformation gives the world the topological aspect of a four-dimensional beer stein. Lines of force could now be drawn on the original sphere which would disappear at P_1 by entering the handle and reissue at P_2. Then P_1 would be a sink and P_2 a source, but no singularity would occur. In similar fashion,

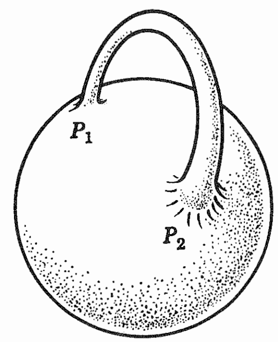


Fig. 15.2

many sources and sinks for the gravitational and electromagnetic fields could be created by an appropriate adjustment of the topology of the world.

Since the topology of a surface is part of its geometry, the field equations would determine the development of geometry and topology in time. That is, we could calculate how the sources and sinks move about in time, and therefore the motion of the particles represented by these sources and sinks would be a consequence of the field equations.

This is an imposing program for the geometrization of classical physics.

It is clear that the actual carrying out of this program will run into great mathematical difficulties, and one cannot yet say that the basic concepts have added anything to our understanding of the physical world. Such ideas may be considered, however, as good examples of the many possibilities hidden in the mathematical structure and geometric concepts which lie at the base of general relativity theory.

Exercises

15.1 An interesting interpretation of the Reissner-Nordstrom metric can be obtained. To see this consider the energy density of the electric field surrounding a point particle. Obtain the total mass-equivalent energy inside a sphere of radius r using the Einstein relation $E = mc^2$, and use Gauss' law to show that its gravitational potential falls off like r^{-2}. If this extra potential is added to the usual point-mass potential in the equation (4.142) for an approximate g_{00}, the result is identical with g_{00} in (15.21).

15.2 Show that for $(\epsilon/M)^2/\kappa < 1/4\pi$ the Reissner-Nordstrom metric has a spherical null surface or one-way membrane, as discussed in Sec. 7.8, while for $(\epsilon/M)^2/\kappa > 1/4\pi$ it does not. What of $(\epsilon/M)^2/\kappa = 1/4\pi$?

15.3 Calculate the Riemann tensor $R^\alpha{}_{\beta\gamma\eta}$ for the Reissner-Nordstrom metric and show that it is singular only at $r = 0$.

15.4 Study the geodesics in the Reissner-Nordstrom metric, in particular the radial null geodesics. How do the geodesics behave in the region of the null surfaces discussed above?

15.5 If $(\epsilon/M)^2/\kappa > 1/4\pi$, the singularity of the Reissner-Nordstrom metric at $r = 0$ is termed *naked* since it is not surrounded by a null surface. Show that light rays, or null geodesics, can pass from the neighborhood of this singularity to large values of r, for example, $r > 2m$, in a finite amount of coordinate time. Hence the region of the singularity is visible to an exterior observer, unlike the situation in the Schwarzschild case.

15.6 Show that the Reissner-Nordstrom metric can be put into the form $\eta_{\mu\nu} - 2ml_\mu l_\nu$ with $l^\mu l_\mu = 0$, as discussed in Chap. 7, by a transformation of the time coordinate (see Exercise 7.11).

15.7 Show that the self-dual Riemann tensor and the self-dual Weyl tensor are equal if $R = 0$ and it is not necessary that $R_{\mu\nu} = 0$. Thus the Petrov classification in the presence of the electromagnetic field can be made with the Riemann tensor, which is simpler than the Weyl tensor (see Exercise 10.8).

15.8 What is the Petrov type of a space-time with a Reissner-Nordstrom metric?

Problems

15.1 A solution of the Einstein-Maxwell equations has been found by Newman and collaborators (1965) that is the generalization of the Reissner-Nordstrom metric in the same sense that the Kerr metric is the generalization of the Schwarzschild metric. It represents the field of a spinning charged body. It is

$$\begin{aligned} ds^2 = {} & \left(1 - \frac{2mr - e^2}{r^2 + a^2\cos^2\theta}\right) c^2\,dt^2 - \frac{r^2 + a^2\cos^2\theta}{\Delta}\,dr^2 \\ & - (r^2 + a^2\cos^2\theta)\,d\theta^2 - \left(r^2 + a^2 + \frac{(2mr - e^2)a^2\sin^2\theta}{r^2 + a^2\cos^2\theta}\right)\sin^2\theta\,d\varphi^2 \\ & - \frac{2a(2mr - e^2)}{r^2 + a^2\cos^2\theta}\sin^2\theta\,d\varphi\,c\,dt \end{aligned}$$

where $\Delta \equiv r^2 - 2mr + e^2 + a^2$. Show that this is indeed a solution of the Einstein-Maxwell equations.

15.2 Study the singularities and the existence of null surfaces in the Kerr-Newman metric introduced above.

15.3 What is the asymptotic magnetic field of the Kerr-Newman metric for large r? What is the effective magnetic moment corresponding to this field? Show that the ratio of magnetic moment to angular momentum is ϵ/M; this is twice the so-called "normal" value that one obtains classically for any distribution of material with a constant ratio of charge to mass density but is the same as the ratio obtained for an electron in Dirac's relativistic quantum theory.

Bibliography

Eddington, A. S. (1923): "The Mathematical Theory of Relativity," Cambridge, England, chap. 7.
Einstein, A. (1955): "The Meaning of Relativity," 5th ed., Princeton, N.J., appendix II.
Fletcher, J. G. (1962): Geometrodynamics, in L. Witten (ed.), "Gravitation: An Introduction to Current Research," New York, pp. 412–437.
Hlavaty, V. (1957): "Geometry of Einstein's Unified Field Theory," Groningen.
London, F. (1927): Quantummechanische Deutung der Theorie von Weyl, *Z. Physik*, **42**:375–389.

Misner, C. W., and J. A. Wheeler (1957): Classical Physics as Geometry: Gravitation, Electromagnetism, Unquantized Charge and Mass as Properties of Curved Empty Space, *Ann. Physics*, **2**:525–603.

Newman, E. T., E. Couch, R. Chinnapared, A. Exton, A. Prakash, and R. Torrence (1965): Metric of a Rotating Charged Mass, *J. Math. Phys.*, **6**:918.

Nordström, G. (1918): On the Energy of the Gravitational Field in Einstein's Theory, *Verhandl. Koninkl. Ned. Akad. Wetenschap., Afdel. Natuurk., Amsterdam*, **26**: 1201–1208.

Penrose, R. (1964): Conformal Treatment of Infinity, in B. DeWitt (ed.), "General Relativity, Groups and Topology," New York.

Rainich, G. Y. (1925): Electrodynamics in the General Relativity Theory, *Trans. Am. Math. Soc.*, **27**:106–130.

Reissner, H. (1916): Über die Eigengravitation des elektrischen Feldes nach der Einstein'schen Theorie, *Ann. Physik*, **50**:106–120.

Schrödinger, E. (1950): "Space-Time Structure," Cambridge, England.

Weyl, H. (1918): Gravitation und Elektrizität, *Sitzber. Preuss. Akad. Wiss. Berlin*, pp. 465–480. Reprinted in Lorentz, Einstein, and Minkowski, "Das Relativitätsprinzip," Leipzig, 1918.

Weyl, H. (1922): "Space, Time, Matter," London, chap. 4, sec. 35.

Wheeler, J. A. (1957): On the Nature of Quantum Geometrodynamics, *Ann. Physics*, **2**:604–614.

Wheeler, J. A. (1961): Geometrodynamics and the Problem of Motion, *Rev. Mod. Phys.*, **33**:63–78.

Wheeler, J. A. (1962): "Geometrodynamics," New York–London.

Index

Index