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

1234567890 KPKP 798765

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

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

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

$$f(x_1,x_2,x_3) = 0 \tag{1}$$

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

$$x_3 = \varphi(x_1,x_2) \tag{2}$$

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

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

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

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

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

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

we have to demand

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

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

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

Now let

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

Then (15) implies clearly

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

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:

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

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

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

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

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

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

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

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

(22′) $$\Phi(\xi_i) = \text{const}$$

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

(23) $$m = \min\Phi \qquad M = \max\Phi$$

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

(23′) $$mr \le \Phi \le Mr \qquad \text{for } \xi_1^2 + \xi_2^2 = r^2$$

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

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

where λ and μ are the roots of the secular equation

(24′) $$\det \|c_{ik} - \lambda \delta_{ik}\| = 0$$

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

(25) $$\xi_i = a_i \cos \lambda t + b_i \sin \lambda t$$

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

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

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 $\Phi = \text{const}$ 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 $F(\xi_i) = \text{const}$ is intersected by any plane through O in an ellipse.

It is now easy to conclude that $F(\xi_i) = \text{const}$ 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

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

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

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

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:

$$\bar{x}^j = f^j(x^0,x^1,x^2,x^3) \qquad (1.1)$$
$$x^k = h^k(\bar{x}^0,\bar{x}^1,\bar{x}^2,\bar{x}^3)$$

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

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

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

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

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

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

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) $$P = (T^{i_1i_2\cdots i_a}_{j_1j_2\cdots j_b})(\xi^{j_1}_{(1)}\xi^{j_2}_{(2)} \cdots \xi^{j_b}_{(b)})(\eta^{(1)}_{i_1}\eta^{(2)}_{i_2} \cdots \eta^{(a)}_{i_a})$$

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:

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

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: $S_{(1)}^{[\alpha\cdots\delta]}$, $S_{(2)}^{[\alpha\cdots\delta]}$, $S_{(3)}^{[\alpha\cdots\delta]}$. 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 $\mathbf{T}$ 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

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

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 $\mathbf{A}$ 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

(1.16) $$T^{\alpha}{}_{\gamma} = g_{\gamma\beta}T^{\alpha\beta}$$
$$T_{\delta\gamma} = g_{\alpha\delta}T^{\alpha}{}_{\gamma} = g_{\alpha\delta}g_{\gamma\beta}T^{\alpha\beta}$$

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

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

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

$$T_\gamma{}^\alpha = g^{\alpha\delta}T_{\gamma\delta}$$
$$T^{\beta\alpha} = g^{\gamma\beta}T_\gamma{}^\alpha = g^{\gamma\beta}g^{\alpha\delta}T_{\gamma\delta} \tag{1.18}$$

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:

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

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

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

(1.20b) 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

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

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 **u**, this becomes

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

or in tensor notation,

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

Since the vector **v** can be chosen arbitrarily, formulas (1.21) and (1.22) imply that

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

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 **u** and **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 **u** and **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 **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 **A** of coefficients a_{ki}.
2. To the set Y_i we apply a linear transformation with a matrix **B** of coefficients b_{ki}.

These transformations give two new sets of quantities:

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

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

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

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

$$(1.25) \qquad \sum_k a_{ki} b_{kj} = \delta_{ij}$$

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

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

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

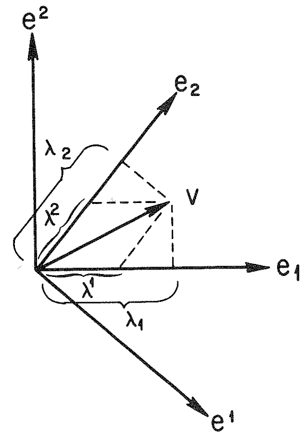
$$\mathbf{A} \cdot \mathbf{B} = \frac{A^i B_i}{\sqrt{A^l A_l B^k B_k}}$$

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

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

where

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