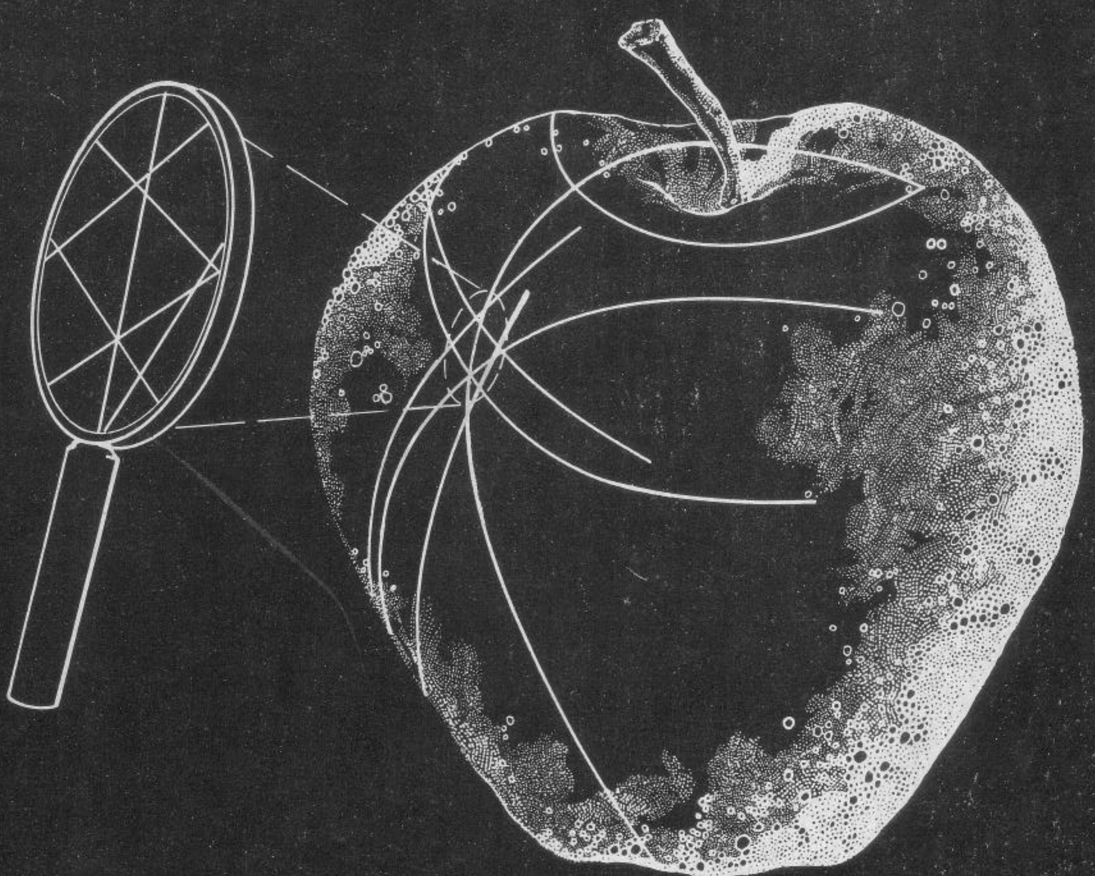
GRAVITATION

Charles W. MISNER Kip S. THORNE John Archibald WHEELER



The Crab Nebula (NGC 1952), the remains of the supernova of July 1054, an event observed and recorded at the Sung national observatory at K'ai-feng. In the intervening 900 years, the debris from the explosion has moved out about three lightyears; i.e., with a speed about 1/300 of that of light. In 1934 Walter Baade and Fritz Zwicky predicted that neutron stars should be produced in supernova explosions. Among the first half-dozen pulsars found in 1968 was one at the center of the Crab Nebula, pulsing 30 times per second, for which there is today no acceptable explanation other than a spinning neutron star. The Chinese historical record shown here lists unusual astronomical phenomena observed during the Northern Sung dynasty. It comes from the "Journal of Astronomy," part 9, chapter 56, of the *Sung History* (*Sung Shih*), first printed in the 1340's. The photograph of that standard record used in this montage is copyright by, and may not be reproduced without permission of, the Trustees of the British Museum.

GRAVITATION

Charles W. MISNER
UNIVERSITY OF MARYLAND

Kip S. THORNE
CALIFORNIA INSTITUTE OF TECHNOLOGY

John Archibald WHEELER
PRINCETON UNIVERSITY

W. H. FREEMAN AND COMPANY San Francisco

Library of Congress Cataloging in Publication Data

Misner, Charles W. 1932–
Gravitation.
Bibliography: p.
1. Gravitation. 2. Astrophysics. 3. General relativity (Physics) I. Thorne, Kip S., 1940– joint author.
II. Wheeler, John Archibald, 1911– joint author.
III. Title.
QC178.M57 531'.14 78-156043
ISBN 0-7167-0334-3
ISBN 0-7167-0344-0 (pbk)

Printed in the United States of America

9 8 7 6

We dedicate this book
To our fellow citizens
Who, for love of truth,
Take from their own wants
By taxes and gifts,
And now and then send forth
One of themselves
As dedicated servant,
To forward the search
Into the mysteries and marvelous simplicities
Of this strange and beautiful Universe,
Our home.

PREFACE

This is a textbook on gravitation physics (Einstein's "general relativity" or "geometrodynamics"). It supplies two tracks through the subject. The first track is focused on the key physical ideas. It assumes, as mathematical prerequisite, only vector analysis and simple partial-differential equations. It is suitable for a one-semester course at the junior or senior level or in graduate school; and it constitutes—in the opinion of the authors—the indispensable core of gravitation theory that every advanced student of physics should learn. The Track-1 material is contained in those pages of the book that have a 1 outlined in gray in the upper outside corner, by which the eye of the reader can quickly pick out the Track-1 sections. In the contents, the same purpose is served by a gray bar beside the section, box, or figure number.

The rest of the text builds up Track 1 into Track 2. Readers and teachers are invited to select, as enrichment material, those portions of Track 2 that interest them most. With a few exceptions, any Track-2 chapter can be understood by readers who have studied only the earlier Track-1 material. The exceptions are spelled out explicitly in "dependency statements" located at the beginning of each Track-2 chapter, or at each transition within a chapter from Track 1 to Track 2.

The entire book (all of Track 1 plus all of Track 2) is designed for a rigorous, full-year course at the graduate level, though many teachers of a full-year course may prefer a more leisurely pace that omits some of the Track-2 material. The full book is intended to give a competence in gravitation physics comparable to that which the average Ph.D. has in electromagnetism. When the student achieves this competence, he knows the laws of physics in flat spacetime (Chapters 1–7). He can predict orders of magnitude. He can also calculate using the principal tools of modern differential geometry (Chapters 8–15), and he can predict at all relevant levels of precision. He understands Einstein's geometric framework for physics (Chapters

16–22). He knows the applications of greatest present-day interest: pulsars and neutron stars (Chapters 23–26); cosmology (Chapters 27–30); the Schwarzschild geometry and gravitational collapse (Chapters 31–34); and gravitational waves (Chapters 35–37). He has probed the experimental tests of Einstein's theory (Chapters 38–40). He will be able to read the modern mathematical literature on differential geometry, and also the latest papers in the physics and astrophysics journals about geometrodynamics and its applications. If he wishes to go beyond the field equations, the four major applications, and the tests, he will find at the end of the book (Chapters 41–44) a brief survey of several advanced topics in general relativity. Among the topics touched on here, superspace and quantum geometrodynamics receive special attention. These chapters identify some of the outstanding physical issues and lines of investigation being pursued today.

Whether the department is physics or astrophysics or mathematics, more students than ever ask for more about general relativity than mere conversation. They want to hear its principal theses clearly stated. They want to know how to "work the handles of its information pump" themselves. More universities than ever respond with a serious course in Einstein's standard 1915 geometrodynamics. What a contrast to Maxwell's standard 1864 electrodynamics! In 1897, when Einstein was a student at Zurich, this subject was not on the instructional calendar of even half the universities of Europe.[1] "We waited in vain for an exposition of Maxwell's theory," says one of Einstein's classmates. "Above all it was Einstein who was disappointed,"[2] for he rated electrodynamics as "the most fascinating subject at the time"[3]—as many students rate Einstein's theory today!

Maxwell's theory recalls Einstein's theory in the time it took to win acceptance. Even as late as 1904 a book could appear by so great an investigator as William Thomson, Lord Kelvin, with the words, "The so-called 'electromagnetic theory of light' has not helped us hitherto . . . it seems to me that it is rather a backward step . . . the one thing about it that seems intelligible to me, I do not think is admissible . . . that there should be an electric displacement perpendicular to the line of propagation."[4] Did the pioneer of the Atlantic cable in the end contribute so richly to Maxwell electrodynamics—from units, and principles of measurement, to the theory of waves guided by wires—because of his own early difficulties with the subject? Then there is hope for many who study Einstein's geometrodynamics today! By the 1920's the weight of developments, from Kelvin's cable to Marconi's wireless, from the atom of Rutherford and Bohr to the new technology of high-frequency circuits, had produced general conviction that Maxwell was right. Doubt dwindled. Confidence led to applications, and applications led to confidence.

Many were slow to take up general relativity in the beginning because it seemed to be poor in applications. Einstein's theory attracts the interest of many today because it is rich in applications. No longer is attention confined to three famous but meager tests: the gravitational red shift, the bending of light by the sun, and

---

[1] G. Holton (1965).
[2] L. Kolbros (1956).
[3] A. Einstein (1949a).
[4] W. Thomson (1904).

Citations for references will be found in the bibliography.

the precession of the perihelion of Mercury around the sun. The combination of radar ranging and general relativity is, step by step, transforming the solar-system celestial mechanics of an older generation to a new subject, with a new level of precision, new kinds of effects, and a new outlook. Pulsars, discovered in 1968, find no acceptable explanation except as the neutron stars predicted in 1934, objects with a central density so high ($\sim 10^{14}$g/cm^3) that the Einstein predictions of mass differ from the Newtonian predictions by 10 to 100 per cent. About further density increase and a final continued gravitational collapse, Newtonian theory is silent. In contrast, Einstein's standard 1915 geometrodynamics predicted in 1939 the properties of a completely collapsed object, a "frozen star" or "black hole." By 1966 detailed digital calculations were available describing the formation of such an object in the collapse of a star with a white-dwarf core. Today hope to discover the first black hole is not least among the forces propelling more than one research: How does rotation influence the properties of a black hole? What kind of pulse of gravitational radiation comes off when such an object is formed? What spectrum of x-rays emerges when gas from a companion star piles up on its way into a black hole?[5] All such investigations and more base themselves on Schwarzschild's standard 1916 static and spherically symmetric solution of Einstein's field equations, first really understood in the modern sense in 1960, and in 1963 generalized to a black hole endowed with angular momentum.

Beyond solar-system tests and applications of relativity, beyond pulsars, neutron stars, and black holes, beyond geometrostatics (compare electrostatics!) and stationary geometries (compare the magnetic field set up by a steady current!) lies geometrodynamics in the full sense of the word (compare electrodynamics!). Nowhere does Einstein's great conception stand out more clearly than here, that the geometry of space is a new physical entity, with degrees of freedom and a dynamics of its own. Deformations in the geometry of space, he predicted in 1918, can transport energy from place to place. Today, thanks to the initiative of Joseph Weber, detectors of such gravitational radiation have been constructed and exploited to give upper limits to the flux of energy streaming past the earth at selected frequencies. Never before has one realized from how many kinds of processes significant gravitational radiation can be anticipated. Never before has there been more interest in picking up this new kind of signal and using it to diagnose faraway events. Never before has there been such a drive in more than one laboratory to raise instrumental sensitivity until gravitational radiation becomes a workaday new window on the universe.

The expansion of the universe is the greatest of all tests of Einstein's geometrodynamics, and cosmology the greatest of all applications. Making a prediction too fantastic for its author to credit, the theory forecast the expansion years before it was observed (1929). Violating the short time-scale that Hubble gave for the expansion, and in the face of "theories" ("steady state"; "continuous creation") manufactured to welcome and utilize this short time-scale, standard general relativity resolutely persisted in the prediction of a long time-scale, decades before the astro-

[5]As of April 1973, there are significant indications that Cygnus X-1 and other compact x-ray sources may be black holes.

physical discovery (1952) that the Hubble scale of distances and times was wrong, and had to be stretched by a factor of more than five. Disagreeing by a factor of the order of thirty with the average density of mass-energy in the universe deduced from astrophysical evidence as recently as 1958, Einstein's theory now as in the past argues for the higher density, proclaims "the mystery of the missing matter," and encourages astrophysics in a continuing search that year by year turns up new indications of matter in the space between the galaxies. General relativity forecast the primordial cosmic fireball radiation, and even an approximate value for its present temperature, seventeen years before the radiation was discovered. This radiation brings information about the universe when it had a thousand times smaller linear dimensions, and a billion times smaller volume, than it does today. Quasistellar objects, discovered in 1963, supply more detailed information from a more recent era, when the universe had a quarter to half its present linear dimensions. Telling about a stage in the evolution of galaxies and the universe reachable in no other way, these objects are more than beacons to light up the far away and long ago. They put out energy at a rate unparalleled anywhere else in the universe. They eject matter with a surprising directivity. They show a puzzling variation with time, different between the microwave and the visible part of the spectrum. Quasistellar objects on a great scale, and galactic nuclei nearer at hand on a smaller scale, voice a challenge to general relativity: help clear up these mysteries!

If its wealth of applications attracts many young astrophysicists to the study of Einstein's geometrodynamics, the same attraction draws those in the world of physics who are concerned with physical cosmology, experimental general relativity, gravitational radiation, and the properties of objects made out of superdense matter. Of quite another motive for study of the subject, to contemplate Einstein's inspiring vision of geometry as the machinery of physics, we shall say nothing here because it speaks out, we hope, in every chapter of this book.

Why a new book? The new applications of general relativity, with their extraordinary physical interest, outdate excellent textbooks of an earlier era, among them even that great treatise on the subject written by Wolfgang Pauli at the age of twenty-one. In addition, differential geometry has undergone a transformation of outlook that isolates the student who is confined in his training to the traditional tensor calculus of the earlier texts. For him it is difficult or impossible either to read the writings of his up-to-date mathematical colleague or to explain the mathematical content of his physical problem to that friendly source of help. We have not seen any way to meet our responsibilities to our students at our three institutions except by a new exposition, aimed at establishing a solid competence in the subject, contemporary in its mathematics, oriented to the physical and astrophysical applications of greatest present-day interest, and animated by belief in the beauty and simplicity of nature.

High Island
South Bristol, Maine
September 4, 1972

Charles W. Misner
Kip S. Thorne
John Archibald Wheeler

CONTENTS

BOXES

FIGURES

ACKNOWLEDGMENTS

Deep appreciation goes to all who made this book possible. A colleague gives us a special lecture so that we may adapt it into one of the chapters of this book. Another investigator clears up for us the tangled history of the production of matter out of the vacuum by strong tidal gravitational forces. A distant colleague telephones in references on the absence of any change in physical constants with time. One student provides a problem on the energy density of a null electromagnetic field. Another supplies curves for effective potential as a function of distance. A librarian writes abroad to get us an article in an obscure publication. A secretary who cares types the third revision of a chapter. Editor and illustrator imaginatively solve a puzzling problem of presentation. Repeat in imagination such instances of warm helpfulness and happy good colleagueship times beyond count. Then one has some impression of the immense debt we owe to over a hundred-fifty colleagues. Each face is etched in our mind, and to each our gratitude is heartfelt. Warm thanks we give also to the California Institute of Technology, the Dublin Institute for Advanced Studies, the Institute for Advanced Study at Princeton, Kyoto University, the University of Maryland, Princeton University, and the University of Texas at Austin for hospitality during the writing of this book. We are grateful to the Academy of Sciences of the U.S.S.R., to Moscow University, and to our Soviet colleagues for their hospitality and the opportunity to become better acquainted in June–July 1971 with Soviet work in gravitation physics. For assistance in the research that went into this book we thank the National Science Foundation for grants (GP27304 and 28027 to Caltech; GP17673 and GP8560 to Maryland; and GP3974 and GP7669 to Princeton); the U.S. Air Force Office of Scientific Research (grant AF49-638-1545 to Princeton); the U.S. National Aeronautics and Space Agency (grant NGR 05-002-256 to Caltech, NSG 210-002-010 to Maryland); the Alfred P. Sloan Foundation for a fellowship awarded to one of us (K.S.T.); and the John Simon Guggenheim Memorial Foundation and All Souls College, Oxford, England, for fellowships awarded to another of us (C.W.M.).

GRAVITATION

PART I

SPACETIME PHYSICS

Wherein the reader is led, once quickly (§ 1.1), then again more slowly, down the highways and a few byways of Einstein's geometrodynamics—without benefit of a good mathematical compass.

CHAPTER 1

GEOMETRODYNAMICS IN BRIEF

§1.1. THE PARABLE OF THE APPLE

One day in the year 1666 Newton had gone to the country, and seeing the fall of an apple, as his niece told me, let himself be led into a deep meditation on the cause which thus draws every object along a line whose extension would pass almost through the center of the Earth.

VOLTAIRE (1738)

Once upon a time a student lay in a garden under an apple tree reflecting on the difference between Einstein's and Newton's views about gravity. He was startled by the fall of an apple nearby. As he looked at the apple, he noticed ants beginning to run along its surface (Figure 1.1). His curiosity aroused, he thought to investigate the principles of navigation followed by an ant. With his magnifying glass, he noted one track carefully, and, taking his knife, made a cut in the apple skin one mm above the track and another cut one mm below it. He peeled off the resulting little highway of skin and laid it out on the face of his book. The track ran as straight as a laser beam along this highway. No more economical path could the ant have found to cover the ten cm from start to end of that strip of skin. Any zigs and zags or even any smooth bend in the path on its way along the apple peel from starting point to end point would have increased its length.

"What a beautiful geodesic," the student commented.

His eye fell on two ants starting off from a common point P in slightly different directions. Their routes happened to carry them through the region of the dimple at the top of the apple, one on each side of it. Each ant conscientiously pursued

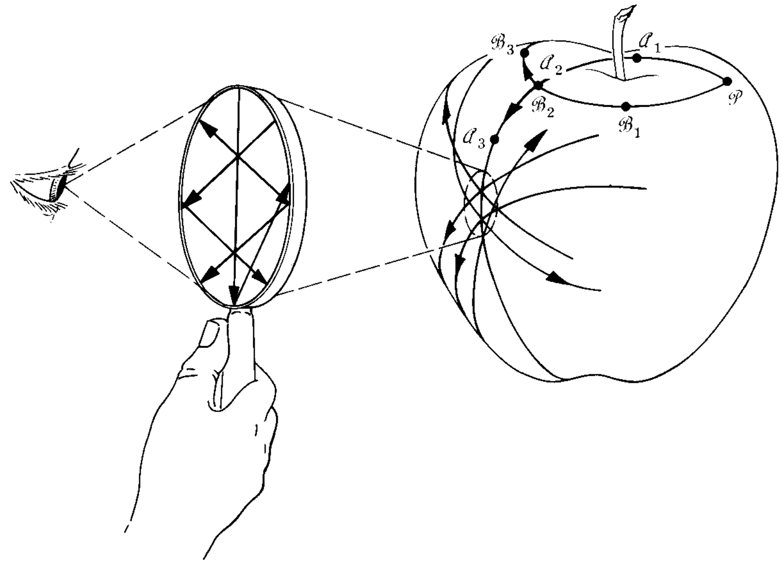


Figure 1.1.
The Riemannian geometry of the spacetime of general relativity is here symbolized by the two-dimensional geometry of the surface of an apple. The geodesic tracks followed by the ants on the apple's surface symbolize the world line followed through spacetime by a free particle. In any sufficiently localized region of spacetime, the geometry can be idealized as flat, as symbolized on the apple's two-dimensional surface by the straight-line course of the tracks viewed in the magnifying glass ("local Lorentz character" of geometry of spacetime). In a region of greater extension, the curvature of the manifold (four-dimensional spacetime in the case of the real physical world; curved two-dimensional geometry in the case of the apple) makes itself felt. Two tracks $\mathcal{A}$ and $\mathcal{B}$, originally diverging from a common point $\mathcal{P}$, later approach, cross, and go off in very different directions. In Newtonian theory this effect is ascribed to gravitation acting at a distance from a center of attraction, symbolized here by the stem of the apple. According to Einstein a particle gets its moving orders locally, from the geometry of spacetime right where it is. Its instructions are simple: to follow the straightest possible track (geodesic). Physics is as simple as it could be locally. Only because spacetime is curved in the large do the tracks cross. Geometrodynamics, in brief, is a double story of the effect of geometry on matter (causing originally divergent geodesics to cross) and the effect of matter on geometry (bending of spacetime initiated by concentration of mass, symbolized by effect of stem on nearby surface of apple).

his geodesic. Each went as straight on his strip of appleskin as he possibly could. Yet because of the curvature of the dimple itself, the two tracks not only crossed but emerged in very different directions.

Einstein's local view of physics contrasted with Newton's "action at a distance"

"What happier illustration of Einstein's geometric theory of gravity could one possibly ask?" murmured the student. "The ants move as if they were attracted by the apple stem. One might have believed in a Newtonian force at a distance. Yet from nowhere does an ant get his moving orders except from the local geometry along his track. This is surely Einstein's concept that all physics takes place by 'local action.' What a difference from Newton's 'action at a distance' view of physics! Now I understand better what this book means."

Physics is simple only when analyzed locally

And so saying, he opened his book and read, "Don't try to describe motion relative to faraway objects. *Physics is simple only when analyzed locally.* And locally the world line that a satellite follows [in spacetime, around the Earth] is already as straight as any world line can be. Forget all this talk about 'deflection' and 'force of gravitation.' I'm inside a spaceship. Or I'm floating outside and near it. Do I feel any 'force of gravitation'? Not at all. Does the spaceship 'feel' such a force? No. Then why talk about it? Recognize that the spaceship and I traverse a region of spacetime free of all force. Acknowledge that the motion through that region is already ideally straight."

The dinner bell was ringing, but still the student sat, musing to himself. "Let me see if I can summarize Einstein's geometric theory of gravity in three ideas: (1) locally, geodesics appear straight; (2) over more extended regions of space and time, geodesics originally receding from each other begin to approach at a rate governed by the curvature of spacetime, and this effect of geometry on matter is what we mean today by that old word 'gravitation'; (3) matter in turn warps geometry. The dimple arises in the apple because the stem is there. I think I see how to put the whole story even more briefly: *Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve.* In other words, matter here," he said, rising and picking up the apple by its stem, "curves space here. To produce a curvature in space here is to force a curvature in space there," he went on, as he watched a lingering ant busily following its geodesic a finger's breadth away from the apple's stem. "Thus matter here influences matter there. That is Einstein's explanation for 'gravitation.'"

Space tells matter how to move

Matter tells space how to curve

Then the dinner bell was quiet, and he was gone, with book, magnifying glass—and apple.

§1.2. SPACETIME WITH AND WITHOUT COORDINATES

Now it came to me: . . . the independence of the gravitational acceleration from the nature of the falling substance, may be expressed as follows: In a gravitational field (of small spatial extension) things behave as they do in a space free of gravitation. . . . This happened in 1908. Why were another seven years required for the construction of the general theory of relativity? The main reason lies in the fact that it is not so easy to free oneself from the idea that coordinates must have an immediate metrical meaning.

ALBERT EINSTEIN [in Schilpp (1949), pp. 65–67.]

Nothing is more distressing on first contact with the idea of "curved spacetime" than the fear that every simple means of measurement has lost its power in this unfamiliar context. One thinks of oneself as confronted with the task of measuring the shape of a gigantic and fantastically sculptured iceberg as one stands with a meter stick in a tossing rowboat on the surface of a heaving ocean. Were it the rowboat itself whose shape were to be measured, the procedure would be simple enough. One would draw it up on shore, turn it upside down, and drive tacks in lightly at strategic points here and there on the surface. The measurement of distances from tack to

Problem: how to measure in curved spacetime

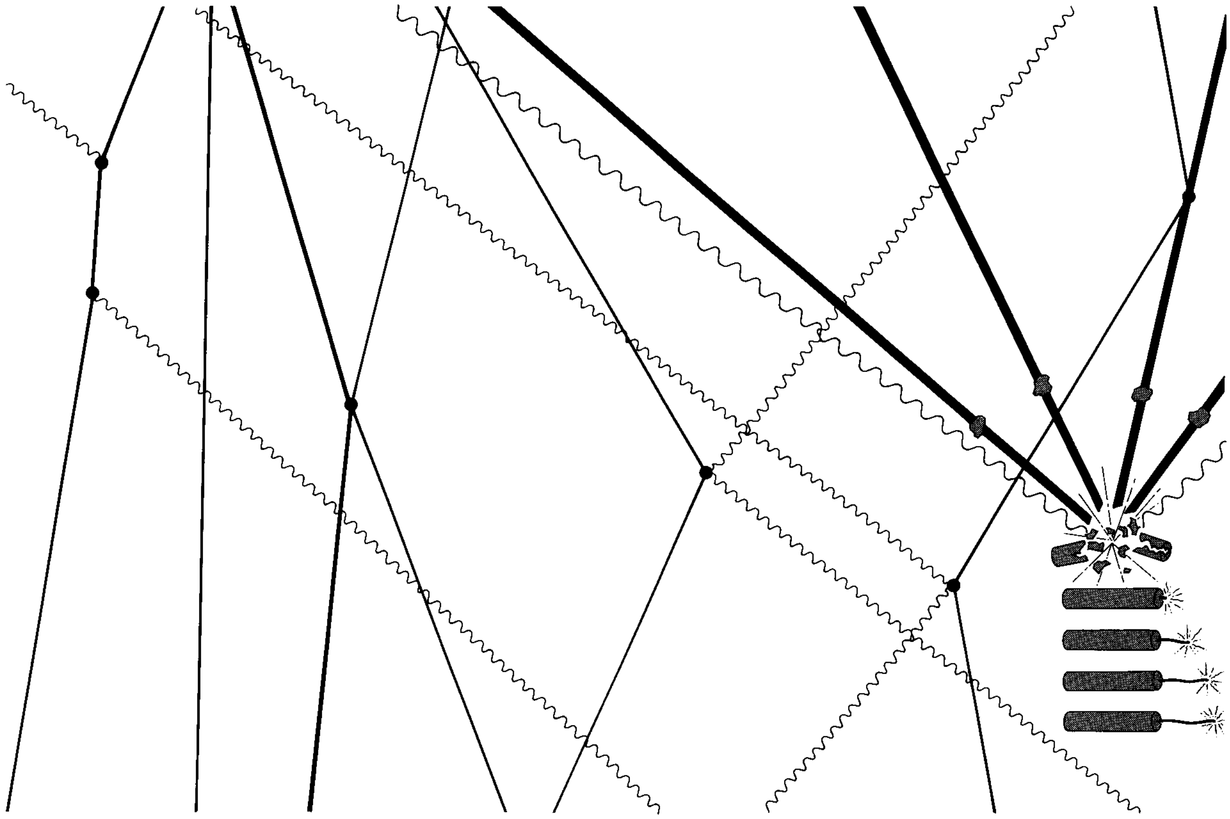


Figure 1.2.
The crossing of straws in a barn full of hay is a symbol for the world lines that fill up spacetime. By their crossings and bends, these world lines mark events with a uniqueness beyond all need of coordinate systems or coordinates. Typical events symbolized in the diagram, from left to right (black dots), are: absorption of a photon; reemission of a photon; collision between a particle and a particle; collision between a photon and a particle; another collision between a photon and a particle; explosion of a firecracker; and collision of a particle from outside with one of the fragments of that firecracker.

Resolution: characterize events by what happens there

tack would record and reveal the shape of the surface. The precision could be made arbitrarily great by making the number of tacks arbitrarily large. It takes more daring to think of driving several score pitons into the towering iceberg. But with all the daring in the world, how is one to drive a nail into spacetime to mark a point? Happily, nature provides its own way to localize a point in spacetime, as Einstein was the first to emphasize. Characterize the point by what happens there! Give a point in spacetime the name "event." Where the event lies is defined as clearly and sharply as where two straws cross each other in a barn full of hay (Figure 1.2). To say that the event marks a collision of such and such a photon with such and such a particle is identification enough. The world lines of that photon and that particle are rooted in the past and stretch out into the future. They have a rich texture of connections with nearby world lines. These nearby world lines in turn are linked in a hundred ways with world lines more remote. How then does one tell the location of an event? Tell first what world lines participate in the event. Next follow each

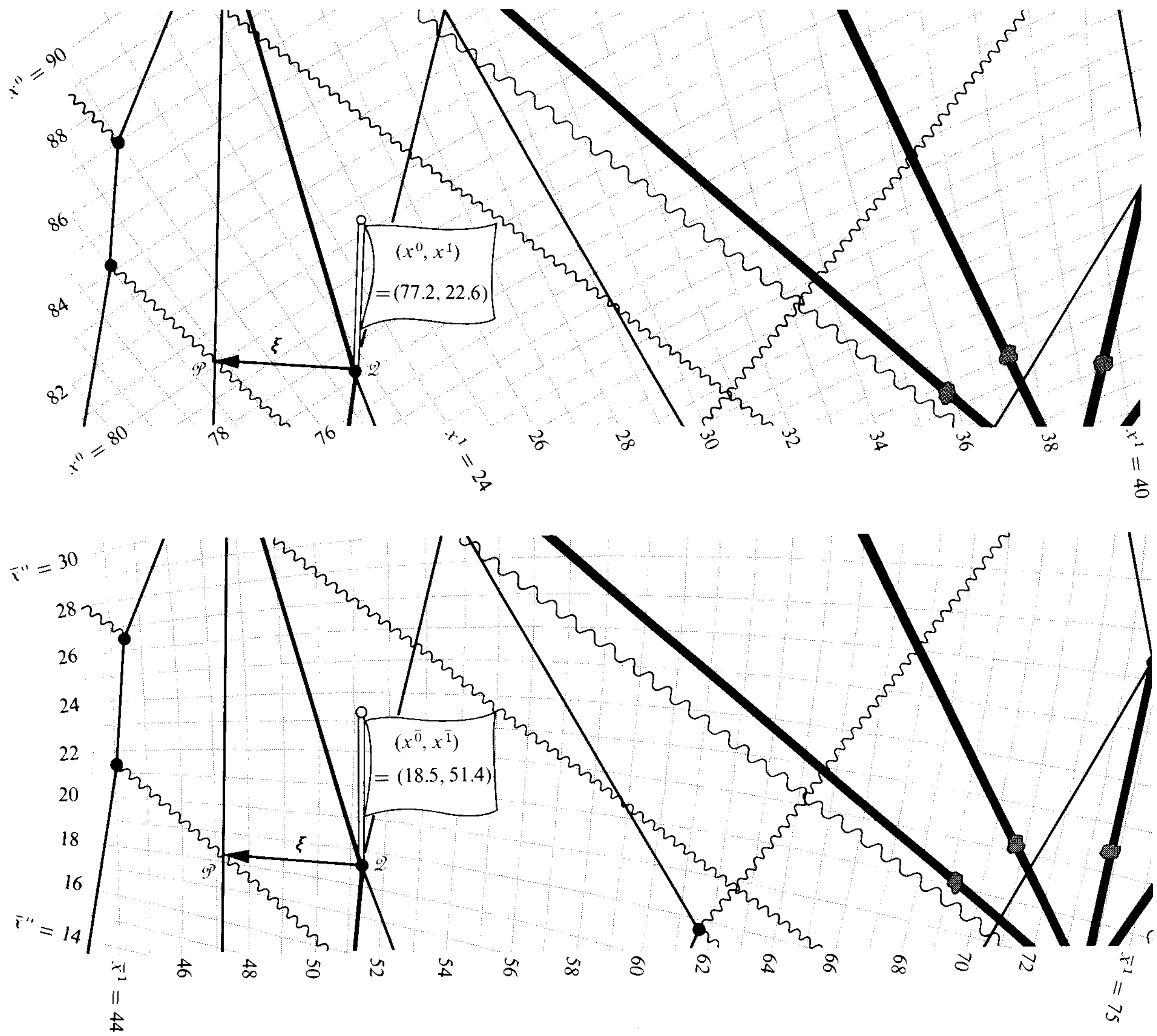


Figure 1.3.
Above: Assigning "telephone numbers" to events by way of a system of coordinates. To say that the coordinate system is "smooth" is to say that events which are almost in the same place have almost the same coordinates. Below: Putting the same set of events into equally good order by way of a different system of coordinates. Picked out specially here are two neighboring events: an event named "$\mathscr{Q}$" with coordinates $(x^0, x^1) = (77.2, 22.6)$ and $(x^{\bar{0}}, x^{\bar{1}}) = (18.5, 51.4)$; and an event named "$\mathscr{P}$" with coordinates $(x^0, x^1) = (79.9, 20.1)$ and $(x^{\bar{0}}, x^{\bar{1}}) = (18.4, 47.1)$. Events $\mathscr{Q}$ and $\mathscr{P}$ are connected by the separation "vector" $\boldsymbol{\xi}$. (Precise definition of a vector in a curved spacetime demands going to the mathematical limit in which the two points have an indefinitely small separation [N-fold reduction of the separation $\mathscr{P} - \mathscr{Q}$], and, in the resultant locally flat space, multiplying the separation up again by the factor N [lim $N \to \infty$; "tangent space"; "tangent vector"]. Forego here that proper way of stating matters, and forego complete accuracy; hence the quote around the word "vector".) In each coordinate system the separation vector $\boldsymbol{\xi}$ is characterized by "components" (differences in coordinate values between $\mathscr{P}$ and $\mathscr{Q}$):

$$(\xi^0, \xi^1) = (79.9 - 77.2,\ 20.1 - 22.6) = (2.7,\ -2.5),$$

$$(\xi^{\bar{0}}, \xi^{\bar{1}}) = (18.4 - 18.5,\ 47.1 - 51.4) = (-0.1,\ -4.3).$$

See Box 1.1 for further discussion of events, coordinates, and vectors.

of these world lines. Name the additional events that they encounter. These events pick out further world lines. Eventually the whole barn of hay is catalogued. Each event is named. One can find one's way as surely to a given intersection as the city dweller can pick his path to the meeting of St. James Street and Piccadilly. No numbers. No coordinate system. No coordinates.

The name of an event can even be arbitrary

That most streets in Japan have no names, and most houses no numbers, illustrates one's ability to do without coordinates. One can abandon the names of two world lines as a means to identify the event where they intersect. Just as one could name a Japanese house after its senior occupant, so one can and often does attach arbitrary names to specific events in spacetime, as in Box 1.1.

Coordinates provide a convenient naming system

Coordinates, however, are convenient. How else from the great thick catalog of events, randomly listed, can one easily discover that along a certain world line one will first encounter event Trinity, then Baker, then Mike, then Argus—but not the same events in some permuted order?

To order events, introduce coordinates! (See Figure 1.3.) Coordinates are four indexed numbers per event in spacetime; on a sheet of paper, only two. Trinity acquires coordinates

$$(x^0, x^1, x^2, x^3) = (77, 23, 64, 11).$$

Coordinates generally do not measure length

In christening events with coordinates, one demands smoothness but foregoes every thought of mensuration. The four numbers for an event are nothing but an elaborate kind of telephone number. Compare their "telephone" numbers to discover whether two events are neighbors. But do not expect to learn how many meters separate them from the difference in their telephone numbers!

Several coordinate systems can be used at once

Nothing prevents a subscriber from being served by competing telephone systems, nor an event from being catalogued by alternative coordinate systems (Figure 1.3). Box 1.1 illustrates the relationships between one coordinate system and another, as well as the notation used to denote coordinates and their transformations.

Vectors

Choose two events, known to be neighbors by the nearness of their coordinate values in a smooth coordinate system. Draw a little arrow from one event to the other. Such an arrow is called a *vector*. (It is a well-defined concept in flat spacetime, or in curved spacetime in the limit of vanishingly small length; for finite lengths in curved spacetime, it must be refined and made precise, under the new name "tangent vector," on which see Chapter 9.) This vector, like events, can be given a name. But whether named "John" or "Charles" or "Kip," it is a unique, well-defined geometrical object. The name is a convenience, but the vector exists even without it.

Just as a quadruple of coordinates

$$(x^0, x^1, x^2, x^3) = (77, 23, 64, 11)$$

is a particularly useful name for the event "Trinity" (it can be used to identify what other events are nearby), so a quadruple of "components"

$$(\xi^0, \xi^1, \xi^2, \xi^3) = (1.2, -0.9, 0, 2.1)$$

Box 1.1 MATHEMATICAL NOTATION FOR EVENTS, COORDINATES, AND VECTORS

Events are denoted by capital script, one-letter Latin names such as Sometimes subscripts are used:	$\mathscr{P}, \mathscr{Q}, \mathscr{A}, \mathscr{B}$. $\mathscr{P}_0, \mathscr{P}_1, \mathscr{B}_6$.
Coordinates of an event $\mathscr{P}$ are denoted by or by or more abstractly by where it is understood that Greek indices can take on any value 0, 1, 2, or 3.	$t(\mathscr{P}), x(\mathscr{P}), y(\mathscr{P}), z(\mathscr{P})$, $x^0(\mathscr{P}), x^1(\mathscr{P}), x^2(\mathscr{P}), x^3(\mathscr{P})$, $x^\mu(\mathscr{P})$ or $x^\alpha(\mathscr{P})$,
Time coordinate (when one of the four is picked to play this role)	$x^0(\mathscr{P})$.
Space coordinates are and are sometimes denoted by It is to be understood that Latin indices take on values 1, 2, or 3.	$x^1(\mathscr{P}), x^2(\mathscr{P}), x^3(\mathscr{P})$ $x^j(\mathscr{P})$ or $x^k(\mathscr{P})$ or....
Shorthand notation: One soon tires of writing explicitly the functional dependence of the coordinates, $x^\beta(\mathscr{P})$; so one adopts the shorthand notation for the coordinates of the event $\mathscr{P}$, and for the space coordinates. One even begins to think of x^β as representing the event $\mathscr{P}$ itself, but must remind oneself that the values of x^0, x^1, x^2, x^3 depend not only on the choice of $\mathscr{P}$ but also on the *arbitrary* choice of coordinates!	x^β x^j
Other coordinates for the same event $\mathscr{P}$ may be denoted EXAMPLE: In Figure 1.3 $(x^0, x^1) = (77.2, 22.6)$ and $(x^{\bar{0}}, x^{\bar{1}}) = (18.5, 51.4)$ refer to the *same* event. The bars, primes, and hats distinguish one coordinate system from another; by putting them on the indices rather than on the x's, we simplify later notation.	$x^{\bar{\alpha}}(\mathscr{P})$ or just $x^{\bar{\alpha}}$, $x^{\alpha'}(\mathscr{P})$ or just $x^{\alpha'}$, $x^{\hat{\alpha}}(\mathscr{P})$ or just $x^{\hat{\alpha}}$.
Transformation from one coordinate system to another is achieved by the four functions which are denoted more succinctly	$x^{\bar{0}}(x^0, x^1, x^2, x^3)$, $x^{\bar{1}}(x^0, x^1, x^2, x^3)$, $x^{\bar{2}}(x^0, x^1, x^2, x^3)$, $x^{\bar{3}}(x^0, x^1, x^2, x^3)$, $x^{\bar{\alpha}}(x^\beta)$.
Separation vector* (little arrow) reaching from one event $\mathscr{Q}$ to neighboring event $\mathscr{P}$ can be denoted abstractly by It can also be characterized by the coordinate-value differences† between $\mathscr{P}$ and $\mathscr{Q}$ (called "components" of the vector)	$\boldsymbol{u}$ or $\boldsymbol{v}$ or $\boldsymbol{\xi}$, or $\mathscr{P} - \mathscr{Q}$. $\xi^\alpha \equiv x^\alpha(\mathscr{P}) - x^\alpha(\mathscr{Q})$, $\xi^{\bar{\alpha}} \equiv x^{\bar{\alpha}}(\mathscr{P}) - x^{\bar{\alpha}}(\mathscr{Q})$.
Transformation of components of a vector from one coordinate system to another is achieved by partial derivatives of transformation equations since $\xi^{\bar{\alpha}} = x^{\bar{\alpha}}(\mathscr{P}) - x^{\bar{\alpha}}(\mathscr{Q}) = (\partial x^{\bar{\alpha}}/\partial x^\beta)[x^\beta(\mathscr{P}) - x^\beta(\mathscr{Q})]$.†	$\xi^{\bar{\alpha}} = \dfrac{\partial x^{\bar{\alpha}}}{\partial x^\beta}\xi^\beta$,
Einstein summation convention is used here: any index that is repeated in a product is automatically summed on	$\dfrac{\partial x^{\bar{\alpha}}}{\partial x^\beta}\xi^\beta \equiv \sum_{\beta=0}^{3}\dfrac{\partial x^{\bar{\alpha}}}{\partial x^\beta}\xi^\beta$.

*This definition of a vector is valid only in flat spacetime. The refined definition ("tangent vector") in curved spacetime is not spelled out here (see Chapter 9), but flat-geometry ideas apply with good approximation even in a curved geometry, when the two points are sufficiently close.

†These formulas are precisely accurate only when the region of spacetime under consideration is flat and when in addition the coordinates are Lorentzian. Otherwise they are approximate—though they become arbitrarily good when the separation between points and the length of the vector become arbitrarily small.

is a convenient name for the vector "John" that reaches from

$$(x^0, x^1, x^2, x^3) = (77, 23, 64, 11)$$

to

$$(x^0, x^1, x^2, x^3) = (78.2, 22.1, 64.0, 13.1).$$

How to work with the components of a vector is explored in Box 1.1.

Coordinate singularities normally unavoidable

There are many ways in which a coordinate system can be imperfect. Figure 1.4 illustrates a coordinate singularity. For another example of a coordinate singularity, run the eye over the surface of a globe to the North Pole. Note the many meridians that meet there ("collapse of cells of egg crates to zero content"). Can't one do better? Find a single coordinate system that will cover the globe without singularity? A theorem says no. Two is the minimum number of "coordinate patches" required to cover the two-sphere without singularity (Figure 1.5). This circumstance emphasizes anew that points and events are primary, whereas coordinates are a mere bookkeeping device.

Continuity of spacetime

Figures 1.2 and 1.3 show only a few world lines and events. A more detailed diagram would show a maze of world lines and of light rays and the intersections between them. From such a picture, one can in imagination step to the idealized limit: an infinitely dense collection of light rays and of world lines of infinitesimal test particles. With this idealized physical limit, the mathematical concept of a continuous four-dimensional "manifold" (four-dimensional space with certain smoothness properties) has a one-to-one correspondence; and in this limit continuous, differentiable (i.e., smooth) coordinate systems operate. The mathematics then supplies a tool to reason about the physics.

The mathematics of manifolds applied to the physics of spacetime

Dimensionality of spacetime

A simple countdown reveals the dimensionality of the manifold. Take a point $\mathscr{P}$ in an n-dimensional manifold. Its neighborhood is an n-dimensional ball (i.e., the interior of a sphere whose surface has $n - 1$ dimensions). Choose this ball so that its boundary is a smooth manifold. The dimensionality of this manifold is $(n - 1)$. In this $(n - 1)$-dimensional manifold, pick a point $\mathscr{Q}$. Its neighborhood is an $(n - 1)$-dimensional ball. Choose this ball so that . . . , and so on. Eventually one comes by this construction to a manifold that is two-dimensional but is not yet known to be two-dimensional (two-sphere). In this two-dimensional manifold, pick a point $\mathscr{M}$. Its neighborhood is a two-dimensional ball ("disc"). Choose this disc so that its boundary is a smooth manifold (circle). In this manifold, pick a point $\mathscr{N}$. Its neighborhood is a one-dimensional ball, but is not yet known to be one-dimensional ("line segment"). The boundaries of this object are two points. This circumstance tells that the intervening manifold is one-dimensional; therefore the previous manifold was two-dimensional; and so on. The dimensionality of the original manifold is equal to the number of points employed in the construction. For spacetime, the dimensionality is 4.

This kind of mathematical reasoning about dimensionality makes good sense at the everyday scale of distances, at atomic distances (10^{-8} cm), at nuclear dimensions (10^{-13} cm), and even at lengths smaller by several powers of ten, if one judges by the concord between prediction and observation in quantum electrodynamics at high

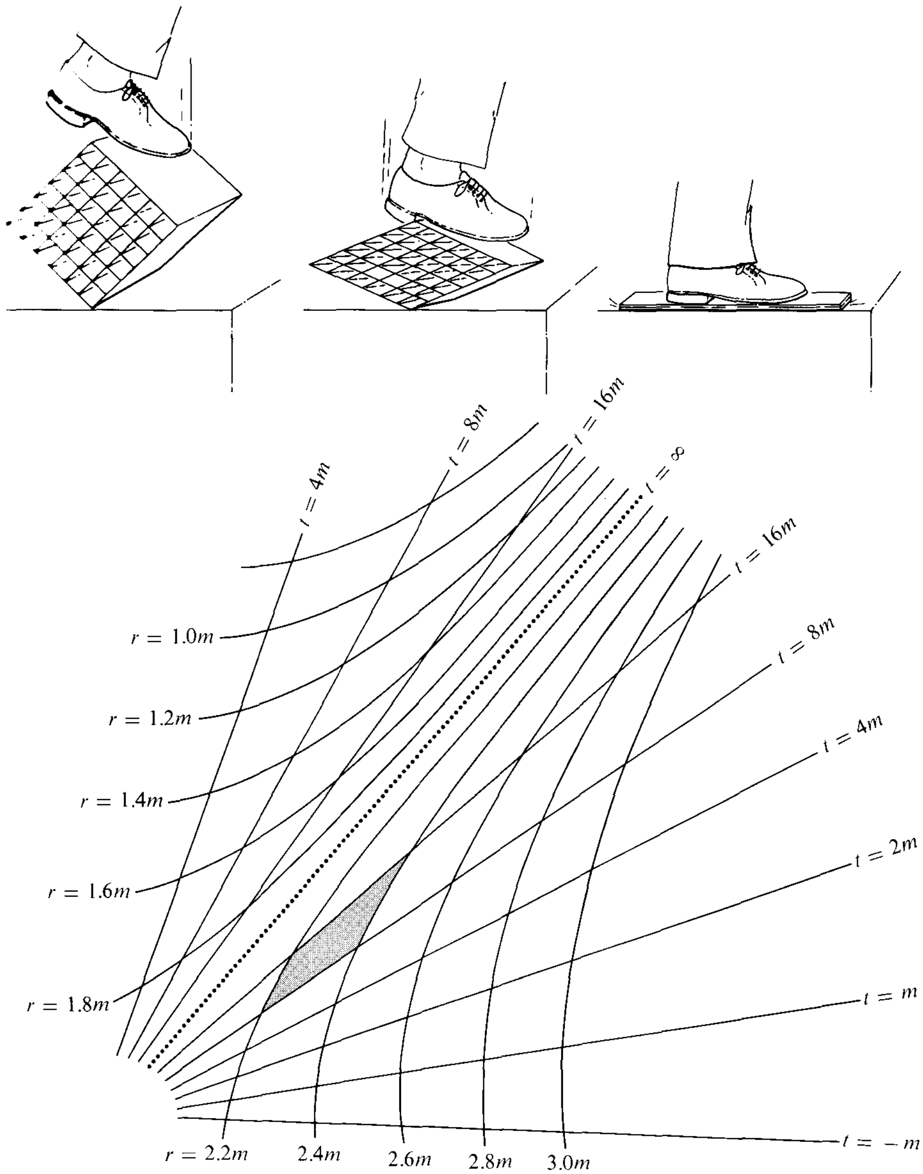


Figure 1.4.

How a mere coordinate singularity arises. Above: A coordinate system becomes *singular* when the "cells in the egg crate" are squashed to zero volume. Below: An example showing such a singularity in the Schwarzschild coordinates r, t often used to describe the geometry around a black hole (Chapter 31). For simplicity the angular coordinates θ, ϕ have been suppressed. The singularity shows itself in two ways. First, all the points along the dotted line, while quite distinct one from another, are designated by the same pair of (r, t) values; namely, $r = 2m$, $t = \infty$. The coordinates provide no way to distinguish these points. Second, the "cells in the egg crate," of which one is shown grey in the diagram, collapse to zero content at the dotted line. In summary, there is nothing strange about the geometry at the dotted line; all the singularity lies in the coordinate system ("poor system of telephone numbers"). No confusion should be permitted to arise from the accidental circumstance that the t coordinate attains an infinite value on the dotted line. No such infinity would occur if t were replaced by the new coordinate $\bar{t}$, defined by

$$(t/2m) = \tan(\bar{t}/2m).$$

When $t = \infty$, the new coordinate $\bar{t}$ is $\bar{t} = \pi m$. The r, $\bar{t}$ coordinates still provide no way to distinguish the points along the dotted line. They still give "cells in the egg crate" collapsed to zero content along the dotted line.

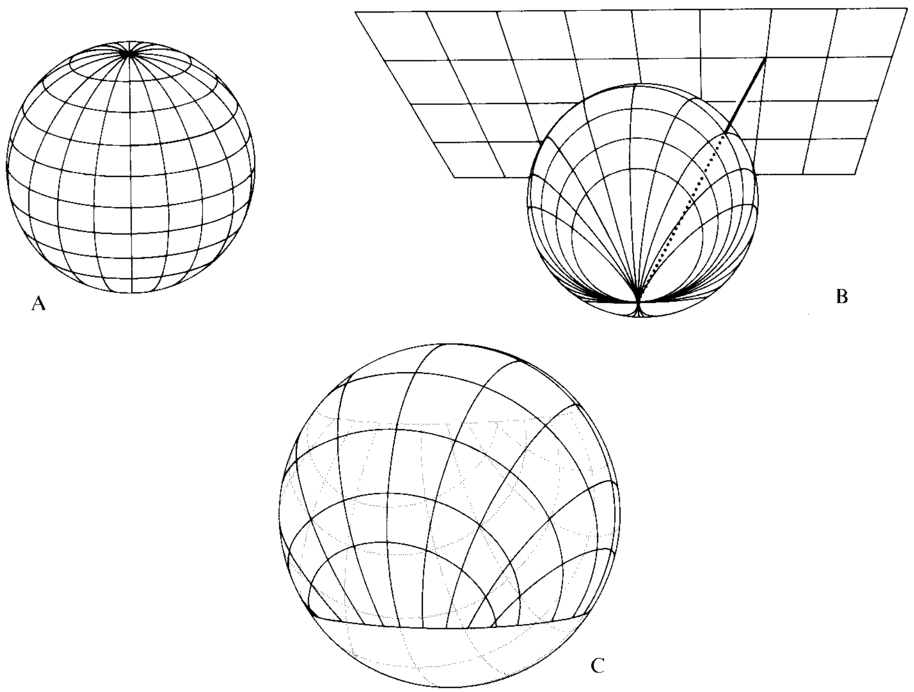


Figure 1.5.
Singularities in familiar coordinates on the two-sphere can be eliminated by covering the sphere with two overlapping coordinate patches. A. Spherical polar coordinates, singular at the North and South Poles, and discontinuous at the international date line. B. Projection of the Euclidean coordinates of the Euclidean two-plane, tangent at the North Pole, onto the sphere via a line running to the South Pole; coordinate singularity at the South Pole. C. Coverage of two-sphere by two overlapping coordinate patches. One, constructed as in B, covers without singularity the northern hemisphere and also the southern tropics down to the Tropic of Capricorn. The other (grey) also covers without singularity all of the tropics and the southern hemisphere besides.

energies (corresponding de Broglie wavelength 10^{-16} cm). Moreover, classical general relativity thinks of the spacetime manifold as a deterministic structure, completely well-defined down to arbitrarily small distances. Not so quantum general relativity or "quantum geometrodynamics." It predicts violent fluctuations in the geometry at distances on the order of the Planck length,

Breakdown in smoothness of spacetime at Planck length

$$\begin{aligned} L^* &= (\hbar G/c^3)^{1/2} \\ &= [(1.054 \times 10^{-27}\text{ g cm}^2/\text{sec})(6.670 \times 10^{-8}\text{ cm}^3/\text{g sec}^2)]^{1/2} \times \\ &\qquad\qquad \times (2.998 \times 10^{10}\text{ cm/sec})^{-3/2} \qquad (1.1) \\ &= 1.616 \times 10^{-33}\text{ cm}. \end{aligned}$$

No one has found any way to escape this prediction. As nearly as one can estimate, these fluctuations give space at small distances a "multiply connected" or "foamlike" character. This lack of smoothness may well deprive even the concept of dimensionality itself of any meaning at the Planck scale of distances. The further exploration of this issue takes one to the frontiers of Einstein's theory (Chapter 44).

If spacetime at small distances is far from the mathematical model of a continuous manifold, is there not also at larger distances a wide gap between the mathematical idealization and the physical reality? The infinitely dense collection of light rays and of world lines of infinitesimal test particles that are to define all the points of the manifold: they surely are beyond practical realization. Nobody has ever found a particle that moves on timelike world lines (finite rest mass) lighter than an electron. A collection of electrons, even if endowed with zero density of charge (e^+ and e^- world lines present in equal numbers) will have a density of mass. This density will curve the very manifold under study. Investigation in infinite detail means unlimited density, and unlimited disturbance of the geometry.

Difficulty in defining geometry even at classical distances?

However, to demand investigatability in infinite detail in the sense just described is as out of place in general relativity as it would be in electrodynamics or gas dynamics. Electrodynamics speaks of the strength of the electric and magnetic field at each point in space and at each moment of time. To measure those fields, it is willing to contemplate infinitesimal test particles scattered everywhere as densely as one pleases. However, the test particles do not have to be there at all to give the field reality. The field has everywhere a clear-cut value and goes about its deterministic dynamic evolution willy-nilly and continuously, infinitesimal test particles or no infinitesimal test particles. Similarly with the geometry of space.

No; one must accept geometry at classical distances as meaningful

In conclusion, when one deals with spacetime in the context of classical physics, one accepts (1) the notion of "infinitesimal test particle" and (2) the idealization that the totality of identifiable events forms a four-dimensional continuous manifold. Only at the end of this book will a look be taken at some of the limitations placed by the quantum principle on one's way of speaking about and analyzing spacetime.

§1.3. WEIGHTLESSNESS

"Gravity is a great mystery. Drop a stone. See it fall. Hear it hit. No one understands why." What a misleading statement! Mystery about fall? What else should the stone do except fall? To fall is normal. The abnormality is an object standing in the way of the stone. If one wishes to pursue a "mystery," do not follow the track of the falling stone. Look instead at the impact, and ask what was the force that pushed the stone away from its natural "world line," (i.e., its natural track through spacetime). That could lead to an interesting issue of solid-state physics, but that is not the topic of concern here. Fall is. Free fall is synonymous with weightlessness: absence of any force to drive the object away from its normal track through spacetime. Travel aboard a freely falling elevator to experience weightlessness. Or travel aboard a spaceship also falling straight toward the Earth. Or, more happily, travel aboard a spaceship in that state of steady fall toward the Earth that marks a circular orbit. In each case one is following a natural track through spacetime.

Free fall is the natural state of motion

The traveler has one chemical composition, the spaceship another; yet they travel together, the traveler weightless in his moving home. Objects of such different nuclear constitution as aluminum and gold fall with accelerations that agree to better than one part in 10^{11}, according to Roll, Krotkov, and Dicke (1964), one of the most important null experiments in all physics (see Figure 1.6). Individual molecules fall in step, too, with macroscopic objects [Estermann, Simpson, and Stern (1938)]; and so do individual neutrons [Dabbs, Harvey, Paya, and Horstmann (1965)], individual

All objects fall with the same acceleration

(*continued on page 16*)

Figure 1.6.
Principle of the Roll-Krotkov-Dicke experiment, which showed that the gravitational accelerations of gold and aluminum are equal to 1 part in 10^{11} or better (Princeton, 1964). In the upper lefthand corner, equal masses of gold and aluminum hang from a supporting bar. This bar in turn is supported at its midpoint. If both objects fall toward the sun with the same acceleration of $g = 0.59$ cm/sec^2, the bar does not turn. If the Au mass receives a higher acceleration, $g + \delta g$, then the gold end of the bar starts to turn toward the sun in the Earth-fixed frame. Twelve hours later the sun is on the other side, pulling the other way. The alternating torque lends itself to recognition against a background of noise because of its precise 24-hour period. Unhappily, any substantial mass nearby, such as an experimenter, located at M, will produce a torque that swamps the effect sought. Therefore the actual arrangement was as shown in the body of the figure. One gold weight and two aluminum weights were supported at the three corners of a horizontal equilateral triangle, 6 cm on a side (three-fold axis of symmetry, giving zero response to all the simplest nonuniformities in the gravitational field). Also, the observers performed all operations remotely to eliminate their own gravitational effects*. To detect a rotation of the torsion balance as small as $\sim 10^{-9}$ rad without disturbing the balance, Roll, Krotkov, and Dicke reflected a very weak light beam from the optically flat back face of the quartz triangle. The image of the source slit fell on a wire of about the same size as the slit image. The light transmitted past the wire fell on a photomultiplier. A separate oscillator circuit drove the wire back and forth across the image at 3,000 hertz. When the image was centered perfectly, only even harmonics of the oscillation frequency appeared in the light intensity. However, when the image was displaced slightly to one side, the fundamental frequency appeared in the light intensity. The electrical output of the photomultiplier then contained a 3,000-hertz component. The magnitude and sign of this component were determined automatically. Equally automatically a proportional D.C. voltage was applied to the electrodes shown in the diagram. It restored the torsion balance to its zero position. The D.C. voltage required to restore the balance to its zero position was recorded as a measure of the torque acting on the pendulum. This torque was Fourier-analyzed over a period of many days. The magnitude of the Fourier component of 24-hour period indicated a ratio $\delta g/g = (0.96 \pm 1.04) \times 10^{-11}$. Aluminum and gold thus fall with the same acceleration, despite their important differences summarized in the table.

Ratios	*Al*	*Au*
$\dfrac{\text{Number of neutrons}}{\text{Number of protons}}$	1.08	1.5
$\dfrac{\text{Mass of kinetic energy of K-electron}}{\text{Rest mass of electron}}$	0.005	0.16
$\dfrac{\text{Electrostatic mass-energy of nucleus}}{\text{Mass of atom}}$	0.001	0.004

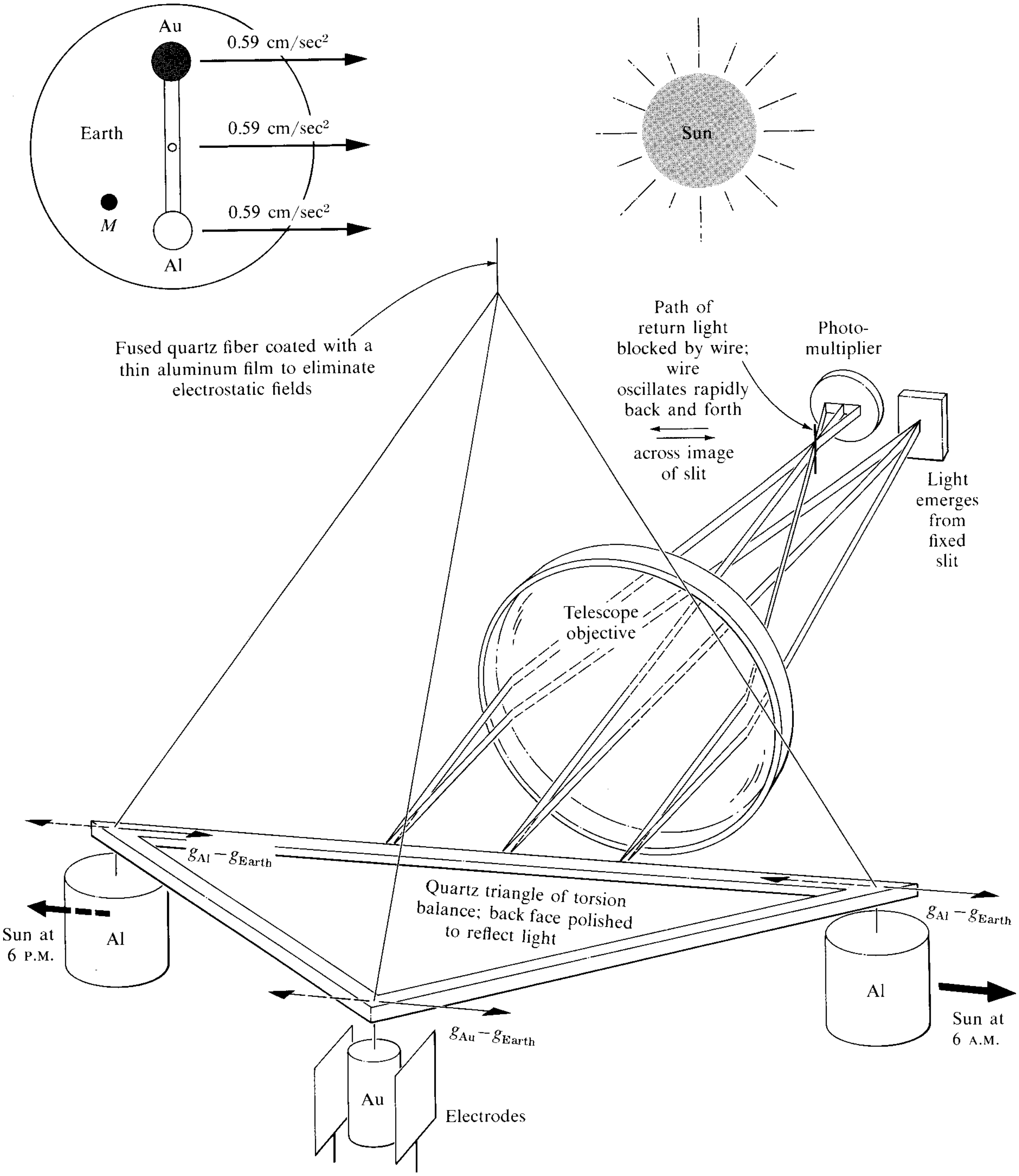
The theoretical implications of this experiment will be discussed in greater detail in Chapters 16 and 38.

Braginsky and Panov (1971) at Moscow University performed an experiment identical in principle to that of Dicke-Roll-Krotkov, but with a modified experimental set-up. Comparing the accelerations of platinum and aluminum rather than of gold and aluminum, they say that

$$\delta g/g \lesssim 1 \times 10^{-12}.$$

---

*Other perturbations had to be, and were, guarded against. (1) A bit of iron on the torsion balance as big as 10^{-3} cm on a side would have contributed, in the Earth's magnetic field, a torque a hundred times greater than the measured torque. (2) The unequal pressure of radiation on the two sides of a mass would have produced an unacceptably large perturbation if the temperature difference between these two sides had exceeded 10^{-4} °K. (3) Gas evolution from one side of a mass would have propelled it like a rocket. If the rate of evolution were as great as 10^{-8} g/day, the calculated force would have been $\sim 10^{-7}$ g cm/sec^2, enough to affect the measurements. (4) The rotation was measured with respect to the pier that supported the equipment. As a guarantee that this pier did not itself rotate, it was anchored to bed rock. (5) Electrostatic forces were eliminated; otherwise they would have perturbed the balance.


Au
0.59 cm/sec^2
Earth
0.59 cm/sec^2
M
0.59 cm/sec^2
Al
Sun
Fused quartz fiber coated with a thin aluminum film to eliminate electrostatic fields
Path of return light blocked by wire; wire oscillates rapidly back and forth across image of slit
Photo-multiplier
Light emerges from fixed slit
Telescope objective
$g_{Al} - g_{Earth}$
Quartz triangle of torsion balance; back face polished to reflect light
$g_{Al} - g_{Earth}$
Sun at 6 P.M.
Al
Al
Sun at 6 A.M.
$g_{Au} - g_{Earth}$
Au
Electrodes

electrons [Witteborn and Fairbank (1967)] and individual mu mesons [Beall (1970)]. What is more, not one of these objects has to see out into space to know how to move.

Contemplate the interior of a spaceship, and a key, penny, nut, and pea by accident or design set free inside. Shielded from all view of the world outside by the walls of the vessel, each object stays at rest relative to the vessel. Or it moves through the room in a straight line with uniform velocity. That is the lesson which experience shouts out.

Forego talk of acceleration! That, paradoxically, is the lesson of the circumstance that "all objects fall with the same acceleration." Whose fault were those accelerations, after all? They came from allowing a groundbased observer into the act. The

Box 1.2 MATERIALS OF THE MOST DIVERSE COMPOSITION FALL WITH THE SAME ACCELERATION ("STANDARD WORLD LINE")

Aristotle: "the downward movement of a mass of gold or lead, or of any other body endowed with weight, is quicker in proportion to its size."

Pre-Galilean literature: metal and wood weights fall at the same rate.

Galileo: (1) "the variation of speed in air between balls of gold, lead, copper, porphyry, and other heavy materials is so slight that in a fall of 100 cubits [about 46 meters] a ball of gold would surely not outstrip one of copper by as much as four fingers. Having observed this, I came to the conclusion that in a medium totally void of resistance all bodies would fall with the same speed." (2) later experiments of greater precision "diluting gravity" and finding same time of descent for different objects along an inclined plane.

Newton: inclined plane replaced by arc of pendulum bob; "time of fall" for bodies of different composition determined by comparing time of oscillation of pendulum bobs of the two materials. Ultimate limit of precision in such experiments limited by problem of determining effective length of each pendulum: (acceleration) = $(2\pi/\text{period})^2$(length).

Lorand von Eötvös, Budapest, 1889 and 1922: compared on the rotating earth the vertical defined by a plumb bob of one material with the vertical defined by a plumb bob of other material. The two hanging masses, by the two unbroken threads that support them, were drawn along identical world lines through spacetime (middle of the laboratory of Eötvös!). If cut free, would they also follow identical tracks through spacetime ("normal world line of test mass")? If so, the acceleration that draws the actual world line from the normal free-fall world line will have a standard value, $\boldsymbol{a}$. The experiment of Eötvös did not try to test agreement on the magnitude of $\boldsymbol{a}$ between the two masses. Doing so would have required (1) cutting the threads and (2) following the fall of the two masses. Eötvös renounced this approach in favor of a static observation that he could make with greater precision, comparing the *direction* of $\boldsymbol{a}$ for the two masses. The direction of the supporting thread, so his argument ran, reveals the direction in which the mass is being dragged away from its normal world line of "free fall" or "weightlessness." This acceleration is the vectorial resultant of (1) an acceleration of magnitude $\boldsymbol{g}$, directed outward against so-called gravity, and (2) an acceleration directed toward the axis of rotation of the earth, of magnitude $\omega^2 R \sin\theta$ (ω, angular vepush of the ground under his feet was driving him away from a natural world line. Through that flaw in his arrangements, he became responsible for all those accelerations. Put him in space and strap rockets to his legs. No difference!* Again the responsibility for what he sees is his. Once more he notes that "all objects fall with

*"No difference" spelled out amounts to Einstein's (1911) principle of the local equivalence between a "gravitational field" and an acceleration: "*We arrive at a very satisfactory interpretation of this law of experience, if we assume that the systems* K *and* K′ *are physically exactly equivalent, that is, if we assume that we may just as well regard the system* K *as being in a space free from gravitational fields, if we then regard* K *as uniformly accelerated. This assumption of exact physical equivalence makes it impossible for us to speak of the absolute acceleration of the system of reference, just as the usual theory of relativity forbids us to talk of the absolute velocity of a system; and it makes the equal falling of all bodies in a gravitational field seem a matter of course.*"

locity; R, radius of earth; θ, polar angle measured from North Pole to location of experiment). This centripetal acceleration has a vertical component $-\omega^2 R \sin^2\theta$ too small to come into discussion. The important component is $\omega^2 R \sin\theta \cos\theta$, directed northward and parallel to the surface of the earth. It deflects the thread by the angle

$$\frac{\text{horizontal acceleration}}{\text{vertical acceleration}} = \frac{\omega^2 R \sin\theta\cos\theta}{g} = \frac{3.4\ \text{cm/sec}^2}{980\ \text{cm/sec}^2}\sin\theta\cos\theta = 1.7\times10^{-3}\ \text{radian at}\ \theta = 45°$$

from the straight line connecting the center of the earth to the point of support. A difference, δg, of one part in 10^8 between g for the two hanging substances would produce a difference in angle of hang of plumb bobs equal to 1.7×10^{-11} radian at Budapest ($\theta = 42.5°$). Eötvös reported $\delta g/g$ less than a few parts in 10^9.

Roll, Krotkov, and Dicke, Princeton, 1964: employed as fiducial acceleration, not the 1.7 cm/sec^2 steady horizontal acceleration, produced by the earth's rotation at $\theta = 45°$, but the daily alternating 0.59 cm/sec^2 produced by the sun's attraction. Reported $|g(\text{Au}) - g(\text{Al})|/g$ less than 1×10^{-11}. See Figure 1.6.

Braginsky and Panov, Moscow, 1971: like Roll, Krotkov, and Dicke, employed Sun's attraction as fiducial acceleration. Reported $|g(\text{Pt}) - g(\text{Al})|/g$ less than 1×10^{-12}.

Beall, 1970: particles that are deflected less by the Earth's or the sun's gravitational field than a photon would be, effectively travel faster than light. If they are charged or have other electromagnetic structure, they would then emit Čerenkov radiation, and reduce their velocity below threshold in less than a micron of travel. The threshold is at energies around 10^3 mc^2. Ultrarelativistic particles in cosmic-ray showers are not easily identified, but observations of 10^{13} eV muons show that muons are not "too light" by as much as 5×10^{-5}. Conversely, a particle P bound more strongly than photons by gravity will transfer the momentum needed to make pair production $\gamma \rightarrow \text{P} + \overline{\text{P}}$ occur within a submicron decay length. The existence of photons with energies above 10^{13} eV shows that $e^\pm$ are not "too heavy" by 5 parts in 10^9, $\mu^\pm$ not by 2 in 10^4, Λ, Ξ^-, Ω^- not by a few per cent.

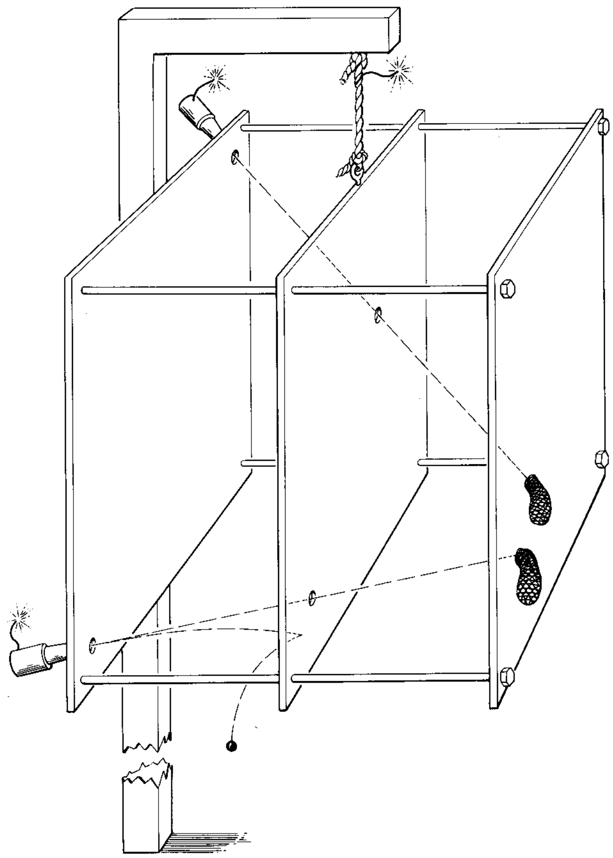


Figure 1.7.
"Weightlessness" as test for a local inertial frame of reference ("Lorentz frame"). Each spring-driven cannon succeeds in driving its projectile, a steel ball bearing, through the aligned holes in the sheets of lucite, and into the woven-mesh pocket, when the frame of reference is free of rotation and in free fall ("normal world line through spacetime"). A cannon would fail (curved and ricocheting trajectory at bottom of drawing) if the frame were hanging as indicated when the cannon went off ("frame drawn away by pull of rope from its normal world line through spacetime"). Harold Waage at Princeton has constructed such a model for an inertial reference frame with lucite sheets about 1 m square. The "fuses" symbolizing time delay were replaced by electric relays. Penetration fails if the frame (1) rotates, (2) accelerates, or (3) does any combination of the two. It is difficult to cite any easily realizable device that more fully illustrates the meaning of the term "local Lorentz frame."

the same acceleration." Physics looks as complicated to the jet-driven observer as it does to the man on the ground. Rule out both observers to make physics look simple. Instead, travel aboard the freely moving spaceship. Nothing could be more natural than what one sees: every free object moves in a straight line with uniform velocity. This is the way to do physics! Work in a very special coordinate system: a coordinate frame in which one is weightless; *a local inertial frame of reference.* Or calculate how things look in such a frame. Or—if one is constrained to a ground-based frame of reference—use a particle moving so fast, and a path length so limited, that the ideal, freely falling frame of reference and the actual ground-based frame get out of alignment by an amount negligible on the scale of the experiment. [Given a 1,500-m linear accelerator, and a 1 GeV electron, time of flight $\simeq (1.5 \times 10^5 \text{ cm})/$

Eliminate the acceleration by use of a local inertial frame

$(3 \times 10^{10}$ cm/sec$) = 0.5 \times 10^{-5}$ sec; fall in this time $\sim\frac{1}{2}gt^2 = (490$ cm/sec$^2)(0.5 \times 10^{-5}$ sec$)^2 \simeq 10^{-8}$ cm.]

In analyzing physics in a local inertial frame of reference, or following an ant on his little section of apple skin, one wins simplicity by foregoing every reference to what is far away. Physics is simple only when viewed locally: that is Einstein's great lesson.

Newton spoke differently: "Absolute space, in its own nature, without relation to anything external, remains always similar and immovable." But how does one give meaning to Newton's absolute space, find its cornerstones, mark out its straight lines? In the real world of gravitation, no particle ever follows one of Newton's straight lines. His ideal geometry is beyond observation. "A comet going past the sun is deviated from an ideal straight line." No. There is no pavement on which to mark out that line. The "ideal straight line" is a myth. It never happened, and it never will.

Newton's absolute space is unobservable, nonexistent

> "It required a severe struggle [for Newton] to arrive at the concept of independent and absolute space, indispensible for the development of theory. . . . Newton's decision was, in the contemporary state of science, the only possible one, and particularly the only fruitful one. But the subsequent development of the problems, proceeding in a roundabout way which no one could then possibly foresee, has shown that the resistance of Leibniz and Huygens, intuitively well-founded but supported by inadequate arguments, was actually justified. . . . It has required no less strenuous exertions subsequently to overcome this concept [of absolute space]"
>
> [A. EINSTEIN (1954)].

What is direct and simple and meaningful, according to Einstein, is the geometry in every local inertial reference frame. There every particle moves in a straight line with uniform velocity. *Define* the local inertial frame so that this simplicity occurs for the first few particles (Figure 1.7). In the frame thus defined, every other free particle is observed also to move in a straight line with uniform velocity. Collision and disintegration processes follow the laws of conservation of momentum and energy of special relativity. That all these miracles come about, as attested by tens of thousands of observations in elementary particle physics, is witness to the inner workings of the machinery of the world. The message is easy to summarize: (1) physics is always and everywhere locally Lorentzian; i.e., locally the laws of special relativity are valid; (2) this simplicity shows most clearly in a local Lorentz frame of reference ("inertial frame of reference"; Figure 1.7); and (3) to test for a local Lorentz frame, test for weightlessness!

But Einstein's local inertial frames exist, are simple

In local inertial frames, physics is Lorentzian

§1.4. LOCAL LORENTZ GEOMETRY, WITH AND WITHOUT COORDINATES

On the surface of an apple within the space of a thumbprint, the geometry is Euclidean (Figure 1.1; the view in the magnifying glass). In spacetime, within a limited region, the geometry is Lorentzian. On the apple the distances between point and point accord with the theorems of Euclid. In spacetime the intervals ("proper distance," "proper time") between event and event satisfy the corresponding theorems of Lorentz-Minkowski geometry (Box 1.3). These theorems lend themselves

Local Lorentz geometry is the spacetime analog of local Euclidean geometry.

(*continued on page 23*)

Box 1.3 LOCAL LORENTZ GEOMETRY AND LOCAL EUCLIDEAN GEOMETRY: WITH AND WITHOUT COORDINATES

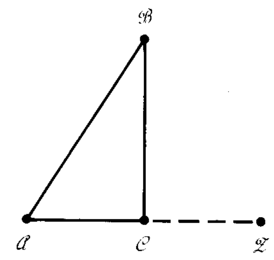
I. Local Euclidean Geometry

What does it mean to say that the geometry of a tiny thumbprint on the apple is Euclidean?

A. *Coordinate-free language* (Euclid):
Given a line $\mathcal{AC}$. Extend it by an equal distance $\mathcal{CZ}$. Let $\mathcal{B}$ be a point not on $\mathcal{AZ}$ but equidistant from $\mathcal{A}$ and $\mathcal{Z}$. Then

$$s_{\mathcal{AB}}^2 = s_{\mathcal{AC}}^2 + s_{\mathcal{BC}}^2.$$

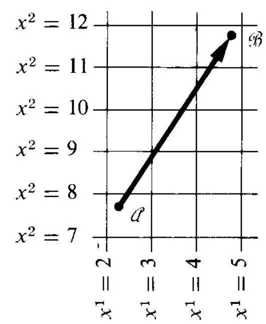
(Theorem of Pythagoras; also other theorems of Euclidean geometry.)



B. *Language of coordinates* (Descartes):
From any point $\mathcal{A}$ to any other point $\mathcal{B}$ there is a distance s given in suitable (Euclidean) coordinates by

$$s_{\mathcal{AB}}^2 = [x^1(\mathcal{B}) - x^1(\mathcal{A})]^2 + [x^2(\mathcal{B}) - x^2(\mathcal{A})]^2.$$

If one succeeds in finding any coordinate system where this is true for all points $\mathcal{A}$ and $\mathcal{B}$ in the thumbprint, then one is guaranteed that (i) this coordinate system is locally Euclidean, and (ii) the geometry of the apple's surface is locally Euclidean.

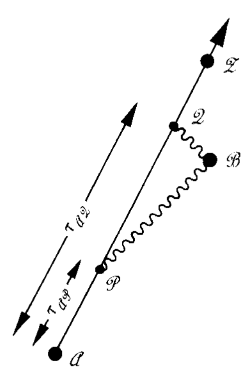


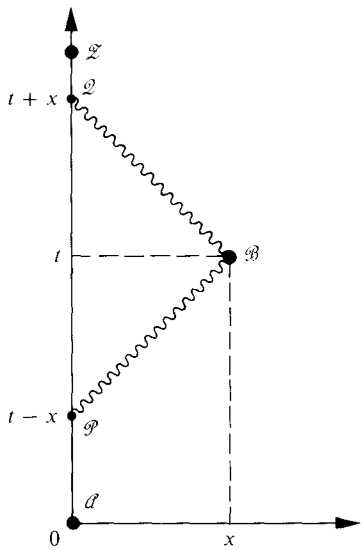
II. Local Lorentz Geometry

What does it mean to say that the geometry of a sufficiently limited region of spacetime in the real physical world is Lorentzian?

A. *Coordinate-free language* (Robb 1936):
Let $\mathcal{AZ}$ be the world line of a free particle. Let $\mathcal{B}$ be an event not on this world line. Let a light ray from $\mathcal{B}$ strike $\mathcal{AZ}$ at the event $\mathcal{Q}$. Let a light ray take off from such an earlier event $\mathcal{P}$ along $\mathcal{AZ}$ that it reaches $\mathcal{B}$. Then the proper distance $s_{\mathcal{AB}}$ (spacelike separation) or proper time $\tau_{\mathcal{AB}}$ (timelike separation) is given by

$$s_{\mathcal{AB}}^2 \equiv -\tau_{\mathcal{AB}}^2 = -\tau_{\mathcal{AQ}}\tau_{\mathcal{AP}}.$$



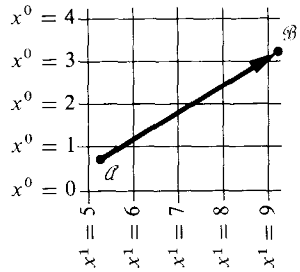


Proof of above criterion for local Lorentz geometry, using coordinate methods in the local Lorentz frame where particle remains at rest:

$$\begin{aligned}\tau_{\mathscr{A}\mathscr{B}}{}^2 &= t^2 - x^2 = (t - x)(t + x)\\ &= \tau_{\mathscr{A}\mathscr{P}}\tau_{\mathscr{A}\mathscr{Q}}.\end{aligned}$$

B. *Language of coordinates* (Lorentz, Poincaré, Minkowski, Einstein):
From any event $\mathscr{A}$ to any other nearby event $\mathscr{B}$, there is a proper distance $s_{\mathscr{A}\mathscr{B}}$ or proper time $\tau_{\mathscr{A}\mathscr{B}}$ given in suitable (local Lorentz) coordinates by

$$\begin{aligned}s_{\mathscr{A}\mathscr{B}}{}^2 = -\tau_{\mathscr{A}\mathscr{B}}{}^2 = &-[x^0(\mathscr{B}) - x^0(\mathscr{A})]^2\\ &+ [x^1(\mathscr{B}) - x^1(\mathscr{A})]^2\\ &+ [x^2(\mathscr{B}) - x^2(\mathscr{A})]^2\\ &+ [x^3(\mathscr{B}) - x^3(\mathscr{A})]^2.\end{aligned}$$



If one succeeds in finding any coordinate system where this is locally true for all neighboring events $\mathscr{A}$ and $\mathscr{B}$, then one is guaranteed that (i) this coordinate system is locally Lorentzian, and (ii) the geometry of spacetime is locally Lorentzian.

III. Statements of Fact

The geometry of an apple's surface is locally Euclidean everywhere. The geometry of spacetime is locally Lorentzian everywhere.

Box 1.3 (*continued*)

IV. Local Geometry in the Language of Modern Mathematics

A. *The metric for any manifold:*
At each point on the apple, at each event of spacetime, indeed, at each point of any "Riemannian manifold," there exists a geometrical object called the *metric tensor* $\boldsymbol{g}$. It is a machine with two input slots for the insertion of two vectors:

$$\begin{array}{c} \text{slot 1} \quad \text{slot 2} \\ \downarrow \qquad \downarrow \\ \boldsymbol{g}(\quad\ ,\quad\ ). \end{array}$$

If one inserts the same vector $\boldsymbol{u}$ into both slots, one gets out the square of the length of $\boldsymbol{u}$:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = \boldsymbol{u}^2.$$

If one inserts two different vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$ (it matters not in which order!), one gets out a number called the "scalar product of $\boldsymbol{u}$ on $\boldsymbol{v}$" and denoted $\boldsymbol{u} \cdot \boldsymbol{v}$:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{u}) = \boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}.$$

The metric is a linear machine:

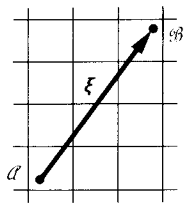
$$\begin{aligned} \boldsymbol{g}(2\boldsymbol{u} + 3\boldsymbol{w}, \boldsymbol{v}) &= 2\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) + 3\boldsymbol{g}(\boldsymbol{w}, \boldsymbol{v}), \\ \boldsymbol{g}(\boldsymbol{u}, a\boldsymbol{v} + b\boldsymbol{w}) &= a\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) + b\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{w}). \end{aligned}$$

Consequently, in a given (arbitrary) coordinate system, its operation on two vectors can be written in terms of their components as a bilinear expression:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) &= g_{\alpha\beta} u^\alpha v^\beta \\ &\quad \text{(implied summation on } \alpha, \beta) \\ &= g_{11} u^1 v^1 + g_{12} u^1 v^2 + g_{21} u^2 v^1 + \cdots. \end{aligned}$$

The quantities $g_{\alpha\beta} = g_{\beta\alpha}$ (α and β running from 0 to 3 in spacetime, from 1 to 2 on the apple) are called the "components of $\boldsymbol{g}$ in the given coordinate system."

B. *Components of the metric in local Lorentz and local Euclidean frames:*
To connect the metric with our previous descriptions of the local geometry, introduce local Euclidean coordinates (on apple) or local Lorentz coordinates (in spacetime).



Let $\boldsymbol{\xi}$ be the separation vector reaching from $\mathcal{A}$ to $\mathcal{B}$. Its components in the local Euclidean (Lorentz) coordinates are

$$\xi^\alpha = x^\alpha(\mathcal{B}) - x^\alpha(\mathcal{A})$$

(cf. Box 1.1). Then the squared length of $\boldsymbol{u}_{\mathcal{A}\mathcal{B}}$, which is the same as the squared distance from $\mathcal{A}$ to $\mathcal{B}$, must be (cf. I.B. and II.B. above)

$$\begin{aligned} \boldsymbol{\xi} \cdot \boldsymbol{\xi} &= \boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\xi}) = g_{\alpha\beta} \xi^\alpha \xi^\beta \\ &= s_{\mathcal{A}\mathcal{B}}{}^2 = (\xi^1)^2 + (\xi^2)^2 \text{ on apple} \\ &= -(\xi^0)^2 + (\xi^1)^2 + (\xi^2)^2 + (\xi^3)^2 \\ &\qquad \text{in spacetime.} \end{aligned}$$

Consequently, the components of the metric are

$$\begin{aligned} &g_{11} = g_{22} = 1,\ g_{12} = g_{21} = 0; \\ &\text{i.e., } g_{\alpha\beta} = \delta_{\alpha\beta} \qquad \text{on apple, in local Euclidean coordinates;} \\ &g_{00} = -1,\ g_{0k} = 0,\ g_{jk} = \delta_{jk} \\ &\qquad \text{in spacetime, in local Lorentz coordinates.} \end{aligned}$$

These special components of the metric in local Lorentz coordinates are written here and hereafter as $g_{\hat{\alpha}\hat{\beta}}$ or $\eta_{\alpha\beta}$, by analogy with the Kronecker delta $\delta_{\alpha\beta}$. In matrix notation:

$$\|g_{\hat{\alpha}\hat{\beta}}\| = \|\eta_{\alpha\beta}\| = \begin{array}{c|cccc} & \multicolumn{4}{c}{\longleftarrow \beta \longrightarrow} \\ & 0 & 1 & 2 & 3 \\ \hline \alpha\ 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 \\ \downarrow\ 3 & 0 & 0 & 0 & 1 \end{array}$$

to empirical test in the appropriate, very special coordinate systems: Euclidean coordinates in Euclidean geometry; the natural generalization of Euclidean coordinates (local Lorentz coordinates; local inertial frame) in the local Lorentz geometry of physics. However, the theorems rise above all coordinate systems in their content. They refer to intervals or distances. Those distances no more call on coordinates for their definition in our day than they did in the time of Euclid. Points in the great pile of hay that is spacetime; and distances between these points: that is geometry! State them in the coordinate-free language or in the language of coordinates: they are the same (Box 1.3).

§1.5. TIME

Time is defined so that motion looks simple.

Time is awake when all things sleep.
Time stands straight when all things fall.
Time shuts in all and will not be shut.
Is, was, and shall be are Time's children.
O Reasoning, be witness, be stable.

VYASA, the *Mahabarata* (*ca.* A.D. 400)

Relative to a local Lorentz frame, a free particle "moves in a straight line with uniform velocity." What "straight" means is clear enough in the model inertial reference frame illustrated in Figure 1.7. But where does the "uniform velocity" come in? Or where does "velocity" show itself? There is not even one clock in the drawing!

A more fully developed model of a Lorentz reference frame will have not only holes, as in Fig. 1.7, but also clock-activated shutters over each hole. The projectile can reach its target only if it (1) travels through the correct region in space and (2) gets through that hole in the correct interval of time ("window in time"). How then is time defined? Time is defined so that motion looks simple!

The time coordinate of a local Lorentz frame is so defined that motion looks simple

No standard of time is more widely used than the day, the time from one high noon to the next. Take that as standard, however, and one will find every good clock or watch clashing with it, for a simple reason. The Earth spins on its axis and also revolves in orbit about the sun. The motion of the sun across the sky arises from neither effect alone, but from the two in combination, different in magnitude though they are. The fast angular velocity of the Earth on its axis (roughly 366.25 complete turns per year) is wonderfully uniform. Not so the apparent angular velocity of the sun about the center of the Earth (one turn per year). It is greater than average by 2 per cent when the Earth in its orbit (eccentricity 0.017) has come 1 per cent closer than average to the sun (Kepler's law) and lower by 2 per cent when the Earth is 1 per cent further than average from the sun. In the first case, the momentary rate of rotation of the sun across the sky, expressed in turns per year, is approximately

$$366.25 - (1 + 0.02);$$

in the other,

$$366.25 - (1 - 0.02).$$

Taking the "mean solar day" to contain $24 \times 3{,}600 = 86{,}400$ standard seconds, one sees that, when the Earth is 1 per cent closer to (or further from) the sun than average, then the number of standard seconds from one high noon to the next is greater (or less) than normal by

$$\frac{0.02 \text{ (drop in turns per year)}}{365.25 \text{ (turns per year on average)}} 86{,}400 \text{ sec} \sim 4.7 \text{ sec.}$$

This is the bookkeeping on time from noon to noon. No standard of time that varies so much from one month to another is acceptable. If adopted, it would make the speed of light vary from month to month!

This lack of uniformity, once recognized (and it was already recognized by the ancients), forces one to abandon the solar day as the standard of time; that day does not make motion look simple. Turn to a new standard that eliminates the motion of the Earth around the sun and concentrates on the spin of the Earth about its axis: the sidereal day, the time between one arrival of a star at the zenith and the next arrival of that star at the zenith. Good! Or good, so long as one's precision of measurement does not allow one to see changes in the intrinsic angular velocity of the Earth. What clock was so bold as first to challenge the spin of the Earth for accuracy? The machinery of the heavens.

Halley (1693) and later others, including Kant (1754), suspected something was amiss from apparent discrepancies between the paths of totality in eclipses of the sun, as predicted by Newtonian gravitation theory using the standard of time then current, and the location of the sites where ancient Greeks and Romans actually recorded an eclipse on the day in question. The moon casts a moving shadow in space. On the day of a solar eclipse, that shadow paints onto the disk of the spinning Earth a black brush stroke, often thousands of kilometers in length, but of width generally much less than a hundred kilometers. He who spins the globe upon the table and wants to make the shadow fall rightly on it must calculate back meticulously to determine two key items: (1) where the moon is relative to Earth and sun at each moment on the ancient day in question; and (2) how much angle the Earth has turned through from then until now. Take the eclipse of Jan. 14, A.D. 484, as an example (Figure 1.8), and assume the same angular velocity for the Earth in the intervening fifteen centuries as the Earth had in 1900 (astronomical reference point). One comes out wrong. The Earth has to be set back by 30° (or the moon moved from its computed position, or some combination of the two effects) to make the Athens observer fall under the black brush. To catch up those 30° (or less, if part of the effect is due to a slow change in the angular momentum of the moon), the Earth had to turn faster in the past than it does today. Assigning most of the discrepancy to terrestrial spin-down (rate of spin-down compatible with modern atomic-clock evidence), and assuming a uniform rate of slowing from then to now

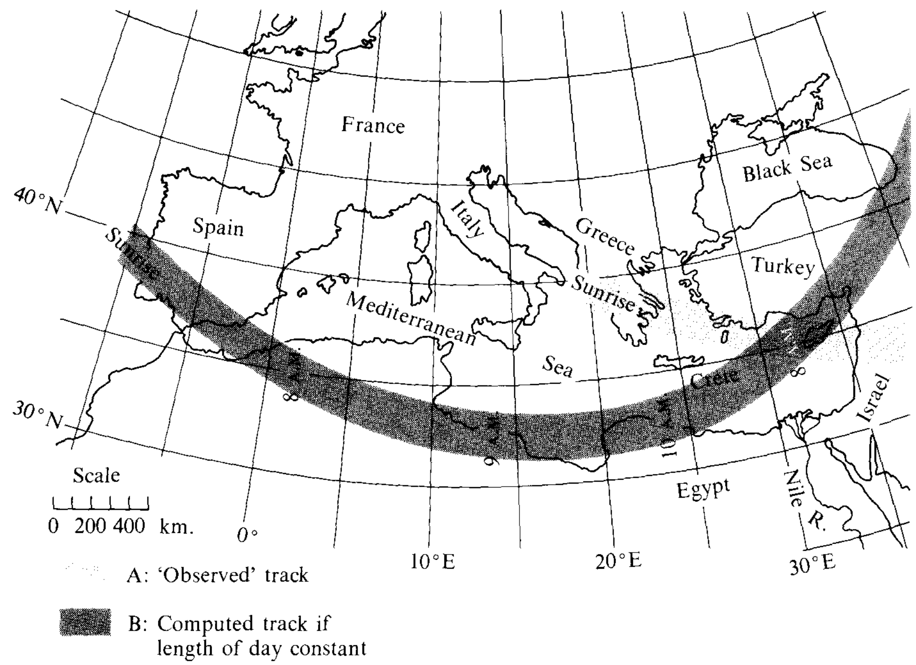


Figure 1.8.
Calculated path of totality for the eclipse of January 14, A.D. 484 (left; calculation based on no spin-down of Earth relative to its 1900 angular velocity) contrasted with the same path as set ahead enough to put the center of totality (at sunrise) at Athens [displacement very close to 30°; actual figure of deceleration adopted in calculations, 32.75 arc sec/(century)^2]. This is "undoubtedly the most reliable of all ancient European eclipses," according to Dr. F. R. Stephenson, of the Department of Geophysics and Planetary Physics of the University of Newcastle upon Tyne, who most kindly prepared this diagram especially for this book. He has also sent a passage from the original Greek biography of Proclus of Athens (died at Athens A.D. 485) by Marinus of Naples, reading, "Nor were there portents wanting in the year which preceded his death; for example, such a great eclipse of the Sun that night seemed to fall by day. For a profound darkness arose so that stars even appeared in the sky. This happened in the eastern sky when the Sun dwelt in Capricorn" [from Westermann and Boissonade (1878)].

Does this 30° for this eclipse, together with corresponding amounts for other eclipses, represent the "right" correction? "Right" is no easy word. From one total eclipse of the sun in the Mediterranean area to another is normally many years. The various provinces of the Greek and Roman worlds were far from having a uniform level of peace and settled life, and even farther from having a uniform standard of what it is to observe an eclipse and put it down for posterity. If the scores of records of the past are unhappily fragmentary, even more unhappy has been the willingness of a few uncritical "investigators" in recent times to rush in and identify this and that historical event with this and that calculated eclipse. Fortunately, by now a great literature is available on the secular deceleration of the Earth's rotation, in the highest tradition of critical scholarship, both astronomical and historical. In addition to the books of O. Neugebauer (1959) and Munk and MacDonald (1960), the paper of Curott (1966), and items cited by these workers, the following are key items. (For direction to them, we thank Professor Otto Neugebauer—no relation to the other Neugebauer cited below!) For the ancient records, and for calculations of the tracks of ancient eclipses, F. K. Ginzel (1882, 1883, 1884); for an atlas of calculated eclipse tracks, Oppolzer (1887) and Ginzel (1899); and for a critical analysis of the evidence. P. V. Neugebauer (1927, 1929, and 1930). This particular eclipse was chosen rather than any other because of the great reliability of the historical record of it.

(angular velocity correction proportional to first power of elapsed time: angle correction itself proportional to square of elapsed time), one estimates from a correction of

$$30° \text{ or } 2 \text{ hours} \qquad 1{,}500 \text{ years ago}$$

the following corrections for intermediate times:

$$30°/10^2, \text{ or } 1.2 \text{ min} \qquad 150 \text{ years ago,}$$

$$30°/10^4, \text{ or } 0.8 \text{ sec} \qquad 15 \text{ years ago.}$$

Thus one sees the downfall of the Earth as a standard of time and its replacement by the orbital motions of the heavenly bodies as a better standard: a standard that does more to "make motion look simple." Astronomical time is itself in turn today being supplanted by atomic time as a standard of reference (see Box 1.4, "Time Today").

Good clocks make spacetime trajectories of free particles look straight

Look at a bad clock for a good view of how time is defined. Let t be time on a "good" clock (time coordinate of a local inertial frame); it makes the tracks of free particles through the local region of spacetime look straight. Let $T(t)$ be the reading of the "bad" clock; it makes the world lines of free particles through the local region of spacetime look curved (Figure 1.9). The old value of the acceleration, translated into the new ("bad") time, becomes

$$0 = \frac{d^2x}{dt^2} = \frac{d}{dt}\left(\frac{dT}{dt}\frac{dx}{dT}\right) = \frac{d^2T}{dt^2}\frac{dx}{dT} + \left(\frac{dT}{dt}\right)^2 \frac{d^2x}{dT^2}.$$

To explain the apparent accelerations of the particles, the user of the new time introduces a force that one knows to be fictitious:

$$F_x = m\frac{d^2x}{dT^2} = -m\frac{\left(\dfrac{dx}{dT}\right)\left(\dfrac{d^2T}{dt^2}\right)}{\left(\dfrac{dT}{dt}\right)^2}. \tag{1.2}$$

It is clear from this example of a "bad" time that Newton thought of a "good" time when he set up the principle that "Time flows uniformly" ($d^2T/dt^2 = 0$). Time is defined to make motion look simple!

The principle of uniformity, taken by itself, leaves free the scale of the time variable. The quantity $T = at + b$ satisfies the requirement as well as t itself. The history of timekeeping discloses many choices of the unit and origin of time. Each one required some human action to give it sanction, from the fiat of a Pharaoh to the communique of a committee. In this book the amount of time it takes light to travel one centimeter is decreed to be the unit of time. Spacelike intervals and timelike intervals are measured in terms of one and the same geometric unit: the centimeter. Any other decision would complicate in analysis what is simple in nature. No other choice would live up to Minkowski's words, "Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

Our choice of unit for measuring time: *the geometrodynamic centimeter.*

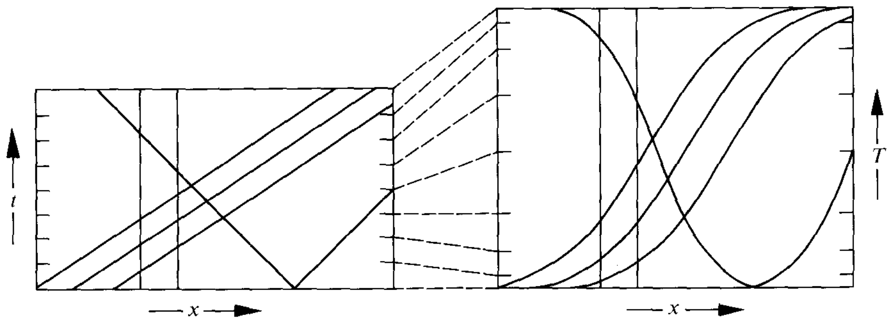


Figure 1.9.
Good clock (left) vs. bad clock (right) as seen in the maps they give of the same free particles moving through the same region of spacetime. The world lines as depicted at the right give the impression that a force is at work. The good definition of time eliminates such fictitious forces. The dashed lines connect corresponding instants on the two time scales.

One can measure time more accurately today than distance. Is that an argument against taking the elementary unit to be the centimeter? No, provided that this definition of the centimeter is accepted: *the geometrodynamic standard centimeter is the fraction*

$$1/(9.460546 \times 10^{17}) \tag{1.3}$$

of the interval between the two "effective equinoxes" that bound the tropical year 1900.0. The tropical year 1900.0 has already been recognized internationally as the fiducial interval by reason of its definiteness and the precision with which it is known. Standards committees have *defined the ephemeris second* so that 31,556,925.974 sec make up that standard interval. Were the speed of light known with perfect precision, the standards committees could have given in the same breath the number of centimeters in the standard interval. But it isn't; it is known to only six decimals. Moreover, the *international centimeter* is defined in terms of the orange-red wavelength of Kr^{86} to only nine decimals (16,507.6373 wavelengths). Yet the standard second is given to 11 decimals. We match the standard second by arbitrarily defining the geometrodynamic standard centimeter so that

$$9.4605460000 \times 10^{17}$$

such centimeters are contained in the standard tropical year 1900.0. The speed of light then becomes exactly

$$\frac{9.4605460000 \times 10^{17}}{31{,}556{,}925.974} \text{ geometrodynamic cm/sec.} \tag{1.4}$$

This is compatible with the speed of light, as known in 1967, in units of "international cm/sec":

$$29{,}979{,}300{,}000 \pm 30{,}000 \text{ international cm/sec.}$$

Box 1.4 TIME TODAY

Prior to 1956 the second was defined as the fraction 1/86,400 of the mean solar day.

From 1956 to 1967 the "second" meant the ephemeris second, defined as the fraction 1/(31,556,925.9747) of the tropical year 00h00m00s December 31, 1899.

Since 1967 the standard second has been the SI (Système International) second, defined as 9,192,631,770 periods of the unperturbed microwave transition between the two hyperfine levels of the ground state of Cs^{133}.

Like the foregoing evolution of the unit for the time *interval*, the evolution of a time *coordinate* has been marked by several stages.

Universal time, UTO, is based on the count of days as they actually occurred historically; in other words, on the actual spin of the earth on its axis; historically, on mean solar time (solar position as corrected by the "equation of time"; i.e., the faster travel of the earth when near the sun than when far from the sun) as determined at Greenwich Observatory.

UT1, the "navigator's time scale," is the same time as corrected for the wobble of the earth on its axis ($\Delta t \sim 0.05$ sec).

UT2 is UT1 as corrected for the periodic fluctuations of unknown origin with periods of one-half year and one year ($\Delta t \sim 0.05$ sec; measured to 3 ms in one day).

Ephemeris Time, ET (as defined by the theory of gravitation and by astronomical observations and calculations), is essentially determined by the orbital motion of the earth around the sun. "Measurement uncertainties limit the realization of accurate ephemeris time to about 0.05 sec for a nine-year average."

Coordinated Universal Time (UTC) is broadcast on stations such as WWV. It was adopted internationally in February 1971 to become effective January 1, 1972. The clock rate is controlled by atomic clocks to be as uniform as possible for one year (atomic time is measured to ~0.1 microsec in 1 min, with diffusion rates of 0.1 microsec per day for ensembles of clocks), but is changed by the infrequent addition or deletion of a second—called a "leap second"—so that UTC never differs more than 0.7 sec from the navigator's time scale, UT1.

THE TIMES
Wednesday
June 21 1972

Time suspended for a second

Time will stand still throughout the world for one second at midnight, June 30. All radio time signals will insert a "leap second" to bring Greenwich Mean Time into line with the earth's loss of three thousandths of a second a day.

The signal from the Royal Greenwich Observatory to Broadcasting House at midnight GMT (1 am BST July 1) will be six short pips marking the seconds 55 to 60 inclusive, followed by a lengthened signal at the following second to mark the new minute.

The foregoing account is abstracted from J. A. Barnes (1971). The following is extracted from a table (not official at time of receipt), kindly supplied by the Time and Frequency Division of the U.S. National Bureau of Standards in Boulder, Colorado.

Timekeeping capabilities of some familiar clocks are as follows:

Tuning fork wrist watch (1960),
1 min/mo.

Quartz crystal clock (1921–1930),
1 μsec/day,
1 sec/yr.

Quartz crystal wrist watch (1971),
0.2 sec/2 mos.,
1 sec/yr.

Cesium beam (atomic resonance, Cs^{133}), (1952–1955),
0.1 μsec/day,
0.5 μsec/mo.

Rubidium gas cell (Rb^{87} resonance), (1957),
0.1 μsec/day,
1–5 μsec/mo.

Hydrogen maser (1960),
0.01 μsec/2 hr,
0.1 μsec/day.

Methane stabilized laser (1969),
0.01 μsec/100 sec.

Recent measurements [Evenson *et al.* (1972)] change the details of the foregoing 1967 argument, but not the principles.

§1.6. CURVATURE

Gravitation seems to have disappeared. Everywhere the geometry of spacetime is locally Lorentzian. And in Lorentz geometry, particles move in a straight line with constant velocity. Where is any gravitational deflection to be seen in that? For answer, turn back to the apple (Figure 1.1). Inspect again the geodesic tracks of the ants on the surface of the apple. Note the reconvergence of two nearby geodesics that originally diverged from a common point. What is the analog in the real world of physics? What analogous concept fits Einstein's injunction that physics is only simple when analyzed locally? Don't look at the distance from the spaceship to the Earth. Look at the distance from the spaceship to a nearby spaceship! Or, to avoid any possible concern about attraction between the two ships, look at two nearby test particles in orbit about the Earth. To avoid distraction by the nonlocal element (the Earth) in the situation, conduct the study in the interior of a spaceship, also in orbit about the Earth. But this region has already been counted as a local inertial frame! What gravitational physics is to be seen there? None. Relative to the spaceship and therefore relative to each other, the two test particles move in a straight line with uniform velocity, to the precision of measurement that is contemplated (see Box 1.5, "Test for Flatness"). Now the key point begins to appear: precision of measurement. Increase it until one begins to discern the gradual acceleration of the test particles away from each other, if they lie along a common radius through the center of the Earth; or toward each other, if their separation lies perpendicular to that line. In Newtonian language, the source of these accelerations is the tide-producing action of the Earth. To the observer in the spaceship, however, no Earth is to be seen. And following Einstein, he knows it is important to analyze motion locally. He represents the separation of the new test particle from the fiducial test particle by the vector ξ^k($k = 1, 2, 3$; components measured in a local Lorentz frame). For the acceleration of this separation, one knows from Newtonian physics what he will find: if the Cartesian z-axis is in the radial direction, then

Gravitation is manifest in relative acceleration of neighboring test particles

$$\begin{aligned} \frac{d^2\xi^x}{dt^2} &= -\frac{Gm_{\text{conv}}}{c^2r^3}\xi^x; \\ \frac{d^2\xi^y}{dt^2} &= -\frac{Gm_{\text{conv}}}{c^2r^3}\xi^y; \\ \frac{d^2\xi^z}{dt^2} &= \frac{2Gm_{\text{conv}}}{c^2r^3}\xi^z. \end{aligned} \tag{1.5}$$

Proof: In Newtonian physics the acceleration of a single particle toward the center of the Earth in conventional units of time is Gm_{conv}/r^2, where G is the Newtonian constant of gravitation, 6.670×10^{-8} cm^3/g sec^2 and m_{conv} is the mass of the Earth in conventional units of grams. In geometric units of time (cm of light-travel time), the acceleration is Gm_{conv}/c^2r^2. When the two particles are separated by a distance ξ perpendicular to r, the one downward acceleration vector is out of line with the other by the angle ξ/r. Consequently one particle accelerates toward the other by the stated amount. When the separation is parallel to r, the relative acceleration is given by evaluating the Newtonian acceleration at r and at $r + \xi$, and taking the difference (ξ times d/dr) Q.E.D. In conclusion, the "local tide-producing acceleration" of Newtonian gravitation theory provides the local description of gravitation that Einstein bids one to seek.

Relative acceleration is caused by curvature

What has this tide-producing acceleration to do with curvature? (See Box 1.6.) Look again at the apple or, better, at a sphere of radius a (Figure 1.10). The separation of nearby geodesics satisfies the "equation of geodesic deviation,"

$$d^2\xi/ds^2 + R\xi = 0. \tag{1.6}$$

Here $R = 1/a^2$ is the so-called Gaussian curvature of the surface. For the surface of the apple, the same equation applies, with the one difference that the curvature R varies from place to place.

Box 1.5 TEST FOR FLATNESS

1. Specify the extension in space L (cm or m) and extension in time T (cm or m of light travel time) of the region under study.

2. Specify the precision $\delta\xi$ with which one can measure the separation of test particles in this region.

3. Follow the motion of test particles moving along initially parallel world lines through this region of spacetime.

4. When the world lines remain parallel to the precision $\delta\xi$ for all directions of travel, then one says that "in a region so limited and to a precision so specified, spacetime is flat."

EXAMPLE: Region just above the surface of the earth, 100 m $\times$ 100 m $\times$ 100 m (space extension), followed for 10^9 m of light-travel time ($T_{\text{conv}} \sim$ 3 sec). Mass of Earth, $m_{\text{conv}} = 5.98 \times 10^{27}$ g, $m = (0.742 \times 10^{-28}$ cm/g$) \times (5.98 \times 10^{27}$ g$) =$ 0.444 cm [see eq. (1.12)]. Tide-producing acceleration $R^z{}_{0z0}$ (relative acceleration in z-direction of two test particles initially at rest and separated from each other by 1 cm of vertical elevation) is

$$\begin{aligned}(d/dr)(m/r^2) &= -2m/r^3 \\ &= -0.888\ \text{cm}/(6.37 \times 10^8\ \text{cm})^3 \\ &= -3.44 \times 10^{-27}\ \text{cm}^{-2}\end{aligned}$$

("cm of relative displacement per cm of light-travel time per cm of light-travel time per cm of vertical separation"). Two test particles with a vertical separation $\xi^z = 10^4$ cm acquire in the time $t = 10^{11}$ cm (difference between time and proper time negligible for such slowly moving test particles) a relative displacement

$$\begin{aligned}\delta\xi^z &= -\tfrac{1}{2}R^z{}_{0z0}t^2\xi^z \\ &= 1.72 \times 10^{-27}\ \text{cm}^{-2}(10^{11}\ \text{cm})^2\ 10^4\ \text{cm} \\ &= 1.72\ \text{mm}.\end{aligned}$$

(Change in relative separation less for other directions of motion). When the minimum uncertainty $\delta\xi$ attainable in a measurement over a 100 m spacing is "worse" than this figure (exceeds 1.72 mm), then to this level of precision the region of spacetime under consideration can be treated as flat. When the uncertainty in measurement is "better" (less) than 1.72 mm, then one must limit attention to a smaller region of space or a shorter interval of time or both, to find a region of spacetime that can be regarded as flat to that precision.

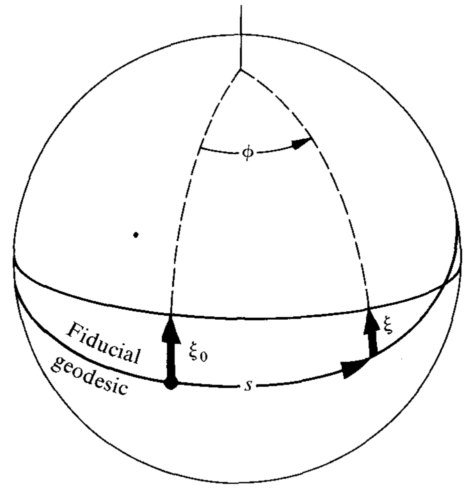


Figure 1.10.
Curvature as manifested in the "acceleration of the separation" of two nearby geodesics. Two geodesics, originally parallel, and separated by the distance ("geodesic deviation") ξ_0, are no longer parallel when followed a distance s. The separation is $\xi = \xi_0 \cos\phi = \xi_0 \cos(s/a)$, where a is the radius of the sphere. The separation follows the equation of simple harmonic motion, $d^2\xi/ds^2 + (1/a^2)\,\xi = 0$ ("equation of geodesic deviation").
The direction of the separation vector, $\boldsymbol{\xi}$, is fixed fully by its orthogonality to the fiducial geodesic. Hence, no reference to the direction of $\boldsymbol{\xi}$ is needed or used in the equation of geodesic deviation; only the magnitude ξ of $\boldsymbol{\xi}$ appears there, and only the magnitude, not direction, of the relative acceleration appears.

In a space of more than two dimensions, an equation of the same general form applies, with several differences. In two dimensions the *direction* of acceleration of one geodesic relative to a nearby, fiducial geodesic is fixed uniquely by the demand that their separation vector, $\boldsymbol{\xi}$, be perpendicular to the fiducial geodesic (see Figure 1.10). Not so in three dimensions or higher. There $\boldsymbol{\xi}$ can remain perpendicular to the fiducial geodesic but rotate about it (Figure 1.11). Thus, to specify the relative acceleration uniquely, one must give not only its magnitude, but also its direction.

The relative acceleration in three dimensions and higher, then, is a vector. Call it "$D^2\boldsymbol{\xi}/ds^2$," and call its four components "$D^2\xi^\alpha/ds^2$." Why the capital D? Why not "$d^2\xi^\alpha/ds^2$"? Because our coordinate system is completely arbitrary (cf. §1.2). The twisting and turning of the coordinate lines can induce changes from point to point in the components ξ^α of $\boldsymbol{\xi}$, even if the vector $\boldsymbol{\xi}$ is not changing at all. Consequently, the accelerations of the components $d^2\xi^\alpha/ds^2$ are generally not equal to the components $D^2\xi^\alpha/ds^2$ of the acceleration!

Curvature is characterized by Riemann tensor

How, then, in curved spacetime can one determine the components $D^2\xi^\alpha/ds^2$ of the relative acceleration? By a more complicated version of the equation of geodesic deviation (1.6). Differential geometry (Part III of this book) provides us with a geometrical object called the *Riemann curvature tensor*, "***Riemann***." ***Riemann*** is

(*continued on page 34*)

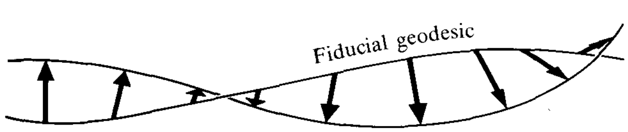
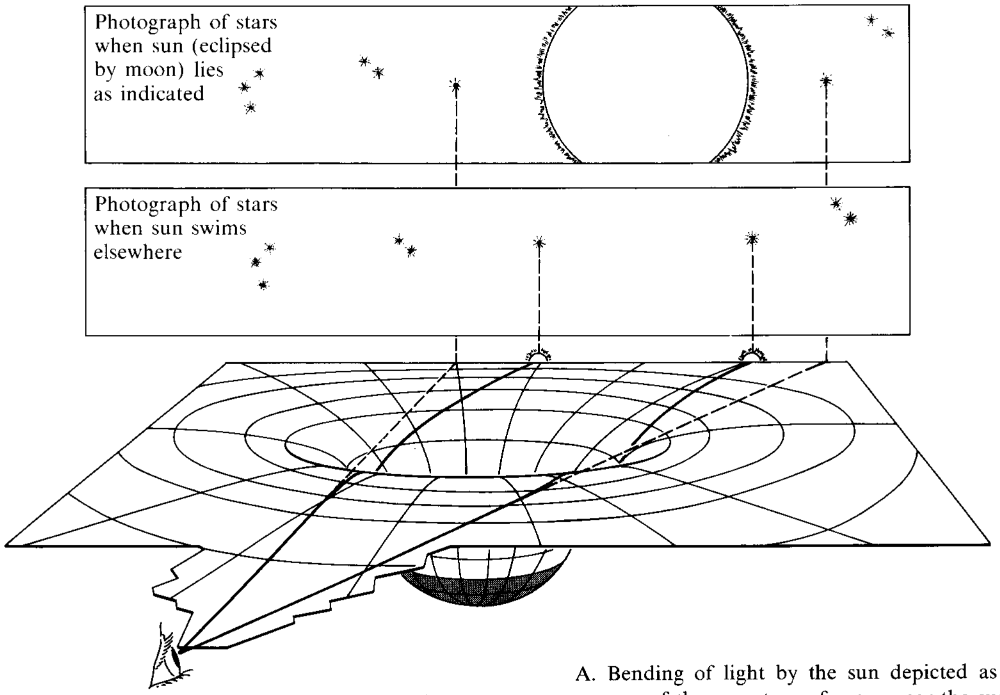


Figure 1.11.
The separation vector $\boldsymbol{\xi}$ between two geodesics in a curved three-dimensional manifold. Here $\boldsymbol{\xi}$ can not only change its length from point to point, but also rotate at a varying rate about the fiducial geodesic. Consequently, the relative acceleration of the geodesics must be characterized by a direction as well as a magnitude; it must be a vector, $D^2\boldsymbol{\xi}/ds^2$.

Box 1.6 CURVATURE OF WHAT?

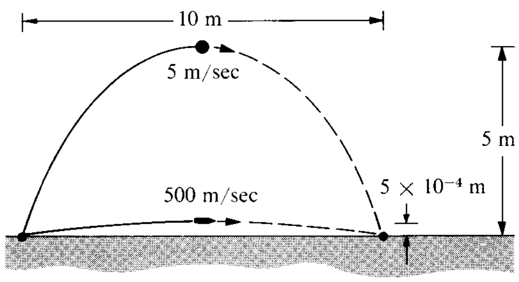
Nothing seems more attractive at first glance than the idea that gravitation is a manifestation of the curvature of space (A), and nothing more ridiculous at a second glance (B). How can the tracks of a ball and of a bullet be curved so differently if that curvature arises from the geometry of space? No wonder that great Riemann did not give the world a geometric theory of gravity. Yes, at the age of 28 (June 10, 1854) he gave the world the mathematical machinery to define and calculate curvature (metric and Riemannian geometry). Yes, he spent his dying days at 40 working to find a unified account of electricity and gravitation. But if there was one reason more than any other why he failed to make the decisive connection between gravitation and curvature, it was this, that he thought of space and the curvature of space, not of spacetime and the curvature of spacetime. To make that forward step took the forty years to special relativity (1905: time on the same footing as space) and then another ten years (1915: general relativity). Depicted in spacetime (C), the tracks of ball and bullet appear to have comparable curvature. In fact, however, neither track has any curvature at all. They both look curved in (C) only because one has forgotten that the spacetime they reside in is itself curved—curved precisely enough to make these tracks the straightest lines in existence ("geodesics").

If it is at first satisfying to see curvature, and curvature of spacetime at that, coming to the fore in so direct a way, then a little more reflection produces a renewed sense of concern. Curvature with respect to what? Not with respect to the labo-

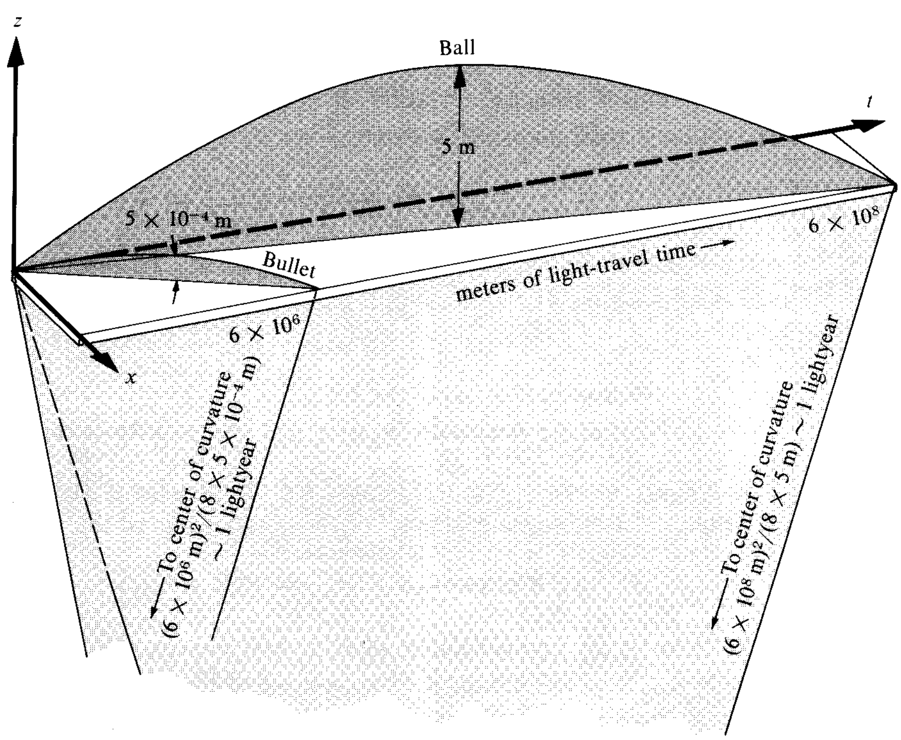


A. Bending of light by the sun depicted as a consequence of the curvature of space near the sun. Ray of light pursues geodesic, but geometry in which it travels is curved (actual travel takes place in spacetime rather than space; correct deflection is twice that given by above elementary picture). Deflection inversely proportional to angular separation between star and center of sun. See Box 40.1 for actual deflections observed at time of an eclipse.

ratory. The earth-bound laboratory has no simple status whatsoever in a proper discussion. First, it is no Lorentz frame. Second, even to mention the earth makes one think of an action-at-a-distance version of gravity (distance from center of earth to ball or bullet). In contrast, it was the whole point of Einstein that physics looks simple only when analyzed locally. To look at local physics, however, means to compare one geodesic of one test particle with geodesics of other test particles traveling (1) nearby with (2) nearly the same directions and (3) nearly the same speeds. Then one can "look at the separations between these nearby test particles and from the second time-rate of change of these separations and the 'equation of geodesic deviation' (equation 1.8) read out the curvature of spacetime."



B. Tracks of ball and bullet through space as seen in laboratory have very different curvatures.



C. Tracks of ball and bullet through spacetime, as recorded in laboratory, have comparable curvatures. Track compared to arc of circle: (radius) = (horizontal distance)^2/8 (rise).

the higher-dimensional analog of the Gaussian curvature R of our apple's surface. ***Riemann*** is the mathematical embodiment of the bends and warps in spacetime. And ***Riemann*** is the agent by which those bends and warps (curvature of spacetime) produce the relative acceleration of geodesics.

Riemann, like the metric tensor $\boldsymbol{g}$ of Box 1.3, can be thought of as a family of machines, one machine residing at each event in spacetime. Each machine has three slots for the insertion of three vectors:

$$\begin{array}{cccc} & \text{slot 1} & \text{slot 2} & \text{slot 3} \\ & \downarrow & \downarrow & \downarrow \\ \textbf{\textit{Riemann}} \; (& \quad , & \quad , & \quad). \end{array}$$

Choose a fiducial geodesic (free-particle world line) passing through an event $\mathcal{Q}$, and denote its unit tangent vector (particle 4-velocity) there by

$$\boldsymbol{u} = \boldsymbol{dx}/d\tau; \text{ components, } u^\alpha = dx^\alpha/d\tau. \tag{1.7}$$

Choose another, neighboring geodesic, and denote by $\boldsymbol{\xi}$ its perpendicular separation from the fiducial geodesic. Then insert $\boldsymbol{u}$ into the first slot of ***Riemann*** at $\mathcal{Q}$, $\boldsymbol{\xi}$ into the second slot, and $\boldsymbol{u}$ into the third. ***Riemann*** will grind for awhile; then out will pop a new vector,

$$\textbf{\textit{Riemann}}\,(\boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{u}).$$

Riemann tensor, through equation of geodesic deviation, produces relative accelerations

The equation of geodesic deviation states that this new vector is the negative of the relative acceleration of the two geodesics:

$$D^2\boldsymbol{\xi}/d\tau^2 + \textbf{\textit{Riemann}}\,(\boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{u}) = 0. \tag{1.8}$$

The Riemann tensor, like the metric tensor (Box 1.3), and like all other tensors, is a linear machine. The vector it puts out is a linear function of each vector inserted into a slot:

$$\begin{aligned} &\textbf{\textit{Riemann}}\,(2\boldsymbol{u}, a\boldsymbol{w} + b\boldsymbol{v}, 3\boldsymbol{r}) \\ &= 2 \times a \times 3\, \textbf{\textit{Riemann}}\,(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{r}) + 2 \times b \times 3\, \textbf{\textit{Riemann}}\,(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{r}). \end{aligned} \tag{1.9}$$

Consequently, in any coordinate system the components of the vector put out can be written as a "trilinear function" of the components of the vectors put in:

$$\boldsymbol{r} = \textbf{\textit{Riemann}}\,(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \Longleftrightarrow r^\alpha = R^\alpha{}_{\beta\gamma\delta}\, u^\beta\, v^\gamma\, w^\delta. \tag{1.10}$$

(Here there is an implied summation on the indices β, γ, δ; cf. Box 1.1.) The $4 \times 4 \times 4 \times 4 = 256$ numbers $R^\alpha{}_{\beta\gamma\delta}$ are called the "components of the Riemann tensor in the given coordinate system." In terms of components, the equation of geodesic deviation states

$$\frac{D^2\xi^\alpha}{d\tau^2} + R^\alpha{}_{\beta\gamma\delta}\frac{dx^\beta}{d\tau}\xi^\gamma\frac{dx^\delta}{d\tau} = 0. \tag{1.8$'$}$$

In Einstein's geometric theory of gravity, this equation of geodesic deviation summarizes the entire effect of geometry on matter. It does for gravitation physics what the Lorentz force equation,

Equation of geodesic deviation is analog of Lorentz force law

$$\frac{D^2x^\alpha}{d\tau^2} - \frac{e}{m} F^\alpha{}_\beta \frac{dx^\alpha}{d\tau} = 0, \tag{1.11}$$

does for electromagnetism. See Box 1.7.

The units of measurement of the curvature are cm^{-2} just as well in spacetime as on the surface of the apple. Nothing does so much to make these units stand out clearly as to express mass in "geometrized units":

Geometrized units

$$\begin{aligned} m(\mathrm{cm}) &= (G/c^2)m_{\mathrm{conv}}(\mathrm{g}) \\ &= (0.742 \times 10^{-28}\ \mathrm{cm/g})m_{\mathrm{conv}}(\mathrm{g}). \end{aligned} \tag{1.12}$$

Box 1.7 EQUATION OF MOTION UNDER THE INFLUENCE OF A GRAVITATIONAL FIELD AND AN ELECTROMAGNETIC FIELD, COMPARED AND CONTRASTED

	Electromagnetism [Lorentz force, equation (1.11)]	*Gravitation [Equation of geodesic deviation (1.8′)]*
Acceleration is defined for one particle?	Yes	No
Acceleration defined how?	Actual world line compared to world line of uncharged "fiducial" test particle passing through same point with same 4-velocity.	Already an uncharged test particle, which can't accelerate relative to itself! Acceleration measured relative to a nearby test particle as fiduciary standard.
Acceleration depends on all four components of the 4-velocity of the particle?	Yes	Yes
Universal acceleration for all test particles in same locations with same 4-velocity?	No; is proportional to e/m	Yes
Driving field	Electromagnetic field	Riemann curvature tensor
Ostensible number of distinct components of driving field	$4 \times 4 = 16$	$4^4 = 256$
Actual number when allowance is made for symmetries of tensor	6	20
Names for more familiar of these components	3 electric 3 magnetic	6 components of local Newtonian tide-producing acceleration

This conversion from grams to centimeters by means of the ratio

$$G/c^2 = 0.742 \times 10^{-28}\ \text{cm/g}$$

is completely analogous to converting from seconds to centimeters by means of the ratio

$$c = \frac{9.4605460000 \times 10^{17}\ \text{cm}}{31{,}556{,}925.974\ \text{sec}}$$

(see end of §1.5). The sun, which in conventional units has $m_{\text{conv}} = 1.989 \times 10^{33}$ g, has in geometrized units a mass $m = 1.477$ km. Box 1.8 gives further discussion.

Using geometrized units, and using the Newtonian theory of gravity, one can readily evaluate nine of the most interesting components of the Riemann curvature tensor near the Earth or the sun. The method is the gravitational analog of determining the electric field strength by measuring the acceleration of a slowly moving test particle. Consider the separation between the geodesics of two nearby and slowly moving ($v \ll c$) particles at a distance r from the Earth or sun. In the standard, nearly inertial coordinates of celestial mechanics, all components of the 4-velocity of the

Components of Riemann tensor evaluated from relative accelerations of slowly moving particles

Box 1.8 GEOMETRIZED UNITS

Throughout this book, we use "geometrized units," in which the speed of light c, Newton's gravitational constant G, and Boltzman's constant k are all equal to unity. The following alternative ways to express the number 1.0 are of great value:

$$1.0 = c = 2.997930 \cdots \times 10^{10}\ \text{cm/sec}$$
$$1.0 = G/c^2 = 0.7425 \times 10^{-28}\ \text{cm/g;}$$
$$1.0 = G/c^4 = 0.826 \times 10^{-49}\ \text{cm/erg;}$$
$$1.0 = Gk/c^4 = 1.140 \times 10^{-65}\ \text{cm/K;}$$
$$1.0 = c^2/G^{1/2} = 3.48 \times 10^{24}\ \text{cm/gauss}^{-1}.$$

One can multiply a factor of unity, expressed in any one of these ways, into any term in any equation without affecting the validity of the equation. Thereby one can convert one's units of measure from grams to centimeters to seconds to ergs to .... For example:

$$\begin{aligned}\text{Mass of sun} = M_\odot &= 1.989 \times 10^{33}\ \text{g}\\ &= (1.989 \times 10^{33}\ g) \times (G/c^2)\\ &= 1.477 \times 10^5\ \text{cm}\\ &= (1.989 \times 10^{33}\ g) \times (c^2)\\ &= 1.788 \times 10^{54}\ \text{ergs.}\end{aligned}$$

The standard unit, in terms of which everything is measured in this book, is centimeters. However, occasionally conventional units are used; in such cases a subscript "conv" is sometimes, but not always, appended to the quantity measured:

$$M_{\odot\text{conv}} = 1.989 \times 10^{33}\ \text{g}.$$

fiducial test particle can be neglected except $dx^0/d\tau = 1$. The space components of the equation of geodesic deviation read

$$d^2\xi^k/d\tau^2 + R^k{}_{0j0}\xi^j = 0. \tag{1.13}$$

Comparing with the conclusions of Newtonian theory, equations (1.5), we arrive at the following information about the curvature of spacetime near a center of mass:

$$\begin{Vmatrix} R^{\hat{x}}{}_{\hat{0}\hat{x}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{x}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{x}\hat{0}} \\ R^{\hat{x}}{}_{\hat{0}\hat{y}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{y}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{y}\hat{0}} \\ R^{\hat{x}}{}_{\hat{0}\hat{z}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{z}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{z}\hat{0}} \end{Vmatrix} \begin{matrix} = \\ = \\ = \end{matrix} \begin{Vmatrix} m/r^3 & 0 & 0 \\ 0 & m/r^3 & 0 \\ 0 & 0 & -2m/r^3 \end{Vmatrix} \tag{1.14}$$

(units cm^{-2}). Here and henceforth the caret or "hat" is used to indicate the components of a vector or tensor in a local Lorentz frame of reference ("physical components," as distinguished from components in a general coordinate system). Einstein's theory will determine the values of the other components of curvature (e.g., $R^{\hat{x}}{}_{\hat{z}\hat{x}\hat{z}} = -m/r^3$); but these nine terms are the ones of principal relevance for many applications of gravitation theory. They are analogous to the components of the electric field in the Lorentz equation of motion. Many of the terms not evaluated are analogous to magnetic field components—ordinarily weak unless the source is in rapid motion.

This ends the survey of the effect of geometry on matter ("effect of curvature of apple in causing geodesics to cross"—especially great near the dimple at the top, just as the curvature of spacetime is especially large near a center of gravitational attraction). Now for the effect of matter on geometry ("effect of stem of apple in causing dimple")!

§1.7. EFFECT OF MATTER ON GEOMETRY

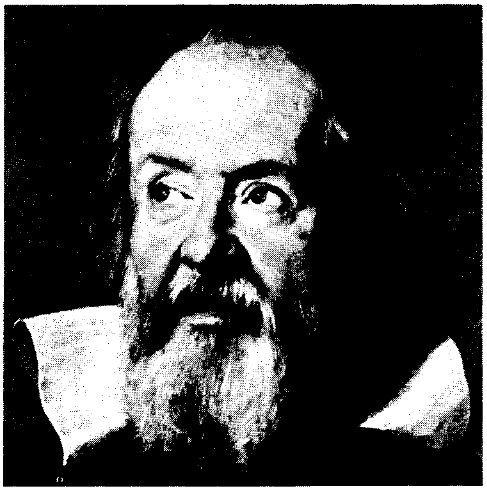
The weight of any heavy body of known weight at a particular distance from the center of the world varies according to the variation of its distance therefrom; so that as often as it is removed from the center, it becomes heavier, and when brought near to it, is lighter. On this account, the relation of gravity to gravity is as the relation of distance to distance from the center.

AL KHĀZINĪ (*Merv, A.D. 1115*), *Book of the Balance of Wisdom*

Figure 1.12 shows a sphere of the same density, $\rho = 5.52$ g/cm^3, as the average density of the Earth. A hole is bored through this sphere. Two test particles, A and B, execute simple harmonic motion in this hole, with an 84-minute period. Therefore their geodesic separation $\boldsymbol{\xi}$, however it may be oriented, undergoes a simple periodic motion with the same 84-minute period:

$$d^2\xi^j/d\tau^2 = -\left(\frac{4\pi}{3}\rho\right)\xi^j, \qquad j = x \text{ or } y \text{ or } z. \tag{1.15}$$

Box 1.9 GALILEO GALILEI
Pisa, February 15, 1564—Arcetri, Florence, January 8, 1642



Uffizi Gallery, Florence

"In questions of science the authority of a thousand is not worth the humble reasoning of a single individual."

GALILEO GALILEI (1632)

"The spaces described by a body falling from rest with a uniformly accelerated motion are to each other as the squares of the time intervals employed in traversing these distances."

GALILEO GALILEI (1638)

"Everything that has been said before and imagined by other people [*about the tides*] *is in my opinion completely invalid. But among the great men who have philosophised about this marvellous effect of nature the one who surprised me the most is Kepler. More than other people he was a person of independent genius, sharp, and had in his hands the motion of the earth. He later pricked up his ears and became interested in the action of the moon on the water, and in other occult phenomena, and similar childishness."*

GALILEO GALILEI (1632)

"It is a most beautiful and delightful sight to behold [*with the new telescope*] *the body of the Moon . . . the Moon certainly does not possess a smooth and polished surface, but one rough and uneven . . . full of vast protuberances, deep chasms and sinuosities . . . stars in myriads, which have never been seen before and which surpass the old, previously known, stars in number more than ten times. I have discovered four planets, neither known nor observed by any one of the astronomers before my time . . . got rid of disputes about the Galaxy or Milky Way, and made its nature clear to the very senses, not to say to the understanding . . . the galaxy is nothing else than a mass of luminous stars planted together in clusters . . . the number of small ones is quite beyond determination—the stars which have been called by every one of the astronomers up to this day* nebulous *are groups of small stars set thick together in a wonderful way."*

GALILEO GALILEI IN *SIDEREUS NUNCIUS* (1610)

"So the principles which are set forth in this treatise will, when taken up by thoughtful minds, lead to many another more remarkable result; and it is to be believed that it will be so on account of the nobility of the subject, which is superior to any other in nature."

GALILEO GALILEI (1638)

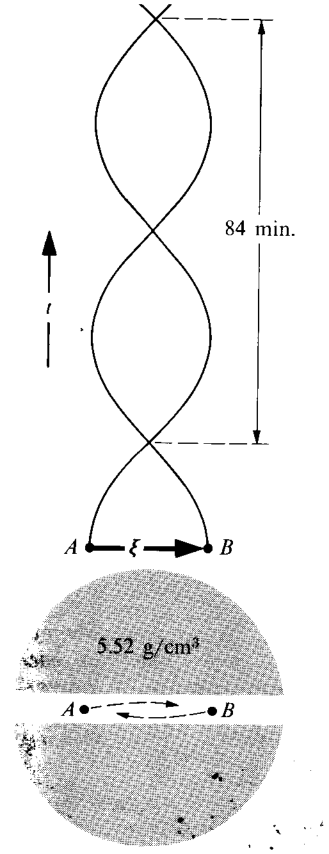


Figure 1.12.
Test particles A and B move up and down a hole bored through the Earth, idealized as of uniform density. At radius r, a particle feels Newtonian acceleration

$$\begin{aligned}\frac{d^2r}{d\tau^2} &= \frac{1}{c^2}\frac{d^2r}{d{t_{\text{conv}}}^2}\\ &= -\frac{G}{c^2}\frac{(\text{mass inside radius } r)}{r^2}\\ &= -\left(\frac{G}{r^2c^2}\right)\left(\frac{4\pi}{3}\rho_{\text{conv}}r^3\right)\\ &= -\omega^2 r.\end{aligned}$$

Consequently, each particle oscillates in simple harmonic motion with precisely the same angular frequency as a satellite, grazing the model Earth, traverses its circular orbit:

$$\omega^2(\text{cm}^{-2}) = \frac{4\pi}{3}\rho(\text{cm}^{-2}),$$

$$\omega^2_{\text{conv}}(\text{sec}^{-2}) = \frac{4\pi G}{3}\rho_{\text{conv}}(\text{g/cm}^3).$$

Comparing this actual motion with the equation of geodesic deviation (1.13) for slowly moving particles in a nearly inertial frame, we can read off some of the curvature components for the interior of this model Earth.

The Riemann tensor inside the Earth

$$\left\|\begin{matrix} R^{\hat x}{}_{\hat 0\hat x\hat 0} & R^{\hat y}{}_{\hat 0\hat x\hat 0} & R^{\hat z}{}_{\hat 0\hat x\hat 0}\\ R^{\hat x}{}_{\hat 0\hat y\hat 0} & R^{\hat y}{}_{\hat 0\hat y\hat 0} & R^{\hat z}{}_{\hat 0\hat y\hat 0}\\ R^{\hat x}{}_{\hat 0\hat z\hat 0} & R^{\hat y}{}_{\hat 0\hat z\hat 0} & R^{\hat z}{}_{\hat 0\hat z\hat 0}\end{matrix}\right\| = (4\pi\rho/3)\left\|\begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix}\right\| \tag{1.16}$$

This example illustrates how the curvature of spacetime is connected to the distribution of matter.

Let a gravitational wave from a supernova pass through the Earth. Idealize the Earth's matter as so nearly incompressible that its density remains practically unchanged. The wave is characterized by ripples in the curvature of spacetime, propagating with the speed of light. The ripples will show up in the components $R^{\hat j}{}_{0k0}$ of the Riemann tensor, and in the relative acceleration of our two test particles. The left side of equation (1.16) will ripple; but the right side will not. Equation (1.16) will break down. No longer will the Riemann curvature be generated directly and solely by the Earth's matter.

Effect of gravitational wave on Riemann tensor

Nevertheless, Einstein tells us, a part of equation (1.16) is undisturbed by the waves: its trace

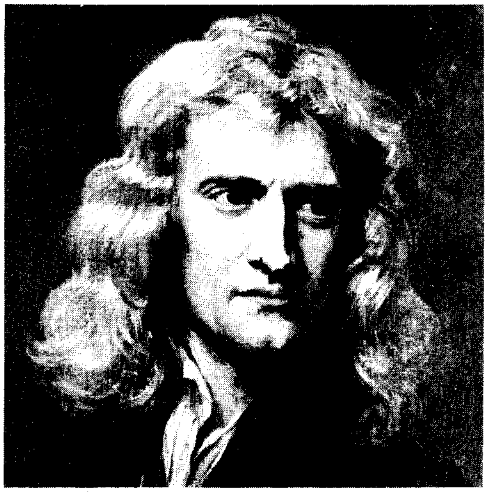
$$R_{\hat{0}\hat{0}} \equiv R^{\hat{x}}{}_{\hat{0}\hat{x}\hat{0}} + R^{\hat{y}}{}_{\hat{0}\hat{y}\hat{0}} + R^{\hat{z}}{}_{\hat{0}\hat{z}\hat{0}} = 4\pi\rho. \tag{1.17}$$

Even in the vacuum outside the Earth this is valid; there both sides vanish [cf. (1.14)].

Einstein tensor introduced

More generally, a certain piece of the Riemann tensor, called the *Einstein tensor* and denoted ***Einstein*** or ***G***, is always generated directly by the local distribution of matter. ***Einstein*** is the geometric object that generalizes $R_{\hat{0}\hat{0}}$, the lefthand side

Box 1.10 ISAAC NEWTON
Woolsthorpe, Lincolnshire, England, December 25, 1642—
Kensington, London, March 20, 1726



"The description of right lines and circles, upon which geometry is founded, belongs to mechanics. Geometry does not teach us to draw these lines, but requires them to be drawn."

[FROM P. 1 OF NEWTON'S PREFACE TO THE FIRST (1687) EDITION OF THE *PRINCIPIA*]

"Absolute space, in its own nature, without relation to anything external, remains always similar and immovable
"Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external."

[FROM THE *SCHOLIUM* IN THE *PRINCIPIA*]

"I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses; for whatever is not reduced from the phenomena is to be called an hypothesis; and hypotheses . . . have no place in experimental philosophy. . . . And to us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea."

[FROM THE *GENERAL SCHOLIUM* ADDED AT THE END OF THE THIRD BOOK OF THE *PRINCIPIA* IN THE SECOND EDITION OF 1713; ESPECIALLY FAMOUS FOR THE PHRASE OFTEN QUOTED FROM NEWTON'S ORIGINAL LATIN, *"HYPOTHESES NON FINGO."*]

"And the same year [1665 or 1666] I began to think of gravity extending to the orb of the Moon, and having found out. . . . All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded Mathematicks and Philosophy more than at any time since."

[FROM MEMORANDUM IN NEWTON'S HANDWRITING ABOUT HIS DISCOVERIES ON FLUXIONS, THE BINOMIAL THEOREM, OPTICS, DYNAMICS, AND GRAVITY, BELIEVED TO HAVE BEEN WRITTEN ABOUT 1714, AND FOUND BY ADAMS ABOUT 1887 IN THE "PORTSMOUTH COLLECTION" OF NEWTON PAPERS] of equation (1.17). Like $R_{\hat{0}\hat{0}}$, ***Einstein*** is a sort of average of ***Riemann*** over all directions. Generating ***Einstein*** and generalizing the righthand side of (1.16) is a geometric object called the *stress-energy tensor* of the matter. It is denoted ***T***. No coordinates are need to define ***Einstein***, and none to define ***T***; like the Riemann tensor, ***Riemann***, and the metric tensor, ***g***, they exist in the complete absence of coordinates. Moreover, in nature they are always equal, aside from a factor of 8π:

Stress-energy tensor introduced

$$\boldsymbol{Einstein} \equiv \boldsymbol{G} = 8\pi \boldsymbol{T}. \tag{1.18}$$

"For hypotheses ought . . . to explain the properties of things and not attempt to predetermine them except in so far as they can be an aid to experiments."

[FROM LETTER OF NEWTON TO I. M. PARDIES, 1672, AS QUOTED IN THE CAJORI NOTES AT THE END OF NEWTON (1687), P. 673]

"That one body may act upon another at a distance through a vacuum, without the mediation of any thing else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man, who has in philosophical matters a competent faculty of thinking, can ever fall into it."

[PASSAGE OFTEN QUOTED BY MICHAEL FARADAY FROM LETTERS OF NEWTON TO RICHARD BENTLY, 1692–1693, AS QUOTED IN THE NOTES OF THE CAJORI EDITION OF NEWTON (1687), P. 643]

"The attractions of gravity, magnetism, and electricity, reach to very sensible distances, and so have been observed . . . ; and there may be others which reach to so small distances as hitherto escape observation; . . . some force, which in immediate contract is exceeding strong, at small distances performs the chemical operations above-mentioned, and reaches not far from the particles with any sensible effect."

[FROM QUERY 31 AT THE END OF NEWTON'S *OPTICKS* (1730)]

"What is there in places almost empty of matter, and whence is it that the sun and planets gravitate towards one another, without dense matter between them? Whence is it that nature doth nothing in vain; and whence arises all that order and beauty which we see in the world? To what end are comets, and whence is it that planets move all one and the same way in orbs concentrick, while comets move all manner of ways in orbs very excentrick; and what hinders the fixed stars from falling upon one another?"

[FROM QUERY 28]

"He is not eternity or infinity, but eternal and infinite; He is not duration or space, but He endures and is present. He endures forever, and is everywhere present; and by existing always and everywhere, He constitutes duration and space. . . . And thus much concerning God; to discourse of whom from the appearances of things, does certainly belong to natural philosophy."

[FROM THE *GENERAL SCHOLIUM* AT THE END OF THE *PRINCIPIA* (1687)]

Einstein field equation: how matter generates curvature

This *Einstein field equation,* rewritten in terms of components in an arbitrary coordinate system, reads

$$G_{\alpha\beta} = 8\pi T_{\alpha\beta}. \tag{1.19}$$

The Einstein field equation is elegant and rich. No equation of physics can be written more simply. And none contains such a treasure of applications and consequences.

Consequences of Einstein field equation

The field equation shows how the stress-energy of matter generates an average curvature (***Einstein*** $\equiv \boldsymbol{G}$) in its neighborhood. Simultaneously, the field equation is a propagation equation for the remaining, anisotropic part of the curvature: it governs the external spacetime curvature of a static source (Earth); it governs the generation of gravitational waves (ripples in curvature of spacetime) by stress-energy in motion; and it governs the propagation of those waves through the universe. The field equation even contains within itself the equations of motion ("Force =

Box 1.11
ALBERT EINSTEIN
Ulm, Germany,
March 14, 1879—
Princeton, New Jersey,
April 18, 1955

Library of E. T. Hochschule, Zürich

Académie des Sciences, Paris

Archives of California Institute of Technology

SEAL: Courtesy of the Lewis and Rosa Strauss Foundation and Princeton University Press mass × acceleration") for the matter whose stress-energy generates the curvature.

Those were some consequences of $\boldsymbol{G} = 8\pi\boldsymbol{T}$. Now for some applications.

Applications of Einstein field equation

The field equation governs the motion of the planets in the solar system; it governs the deflection of light by the sun; it governs the collapse of a star to form a black hole; it determines uniquely the external spacetime geometry of a black hole ("a black hole has no hair"); it governs the evolution of spacetime singularities at the end point of collapse; it governs the expansion and recontraction of the universe. And more; much more.

In order to understand how the simple equation $\boldsymbol{G} = 8\pi\boldsymbol{T}$ can be so all powerful, it is desirable to backtrack, and spend a few chapters rebuilding the entire picture of spacetime, of its curvature, and of its laws, this time with greater care, detail, and mathematics.

Thus ends this survey of the effect of geometry on matter, and the reaction of matter back on geometry, rounding out the parable of the apple.

"What really interests me is whether God had any choice in the creation of the world"

EINSTEIN TO AN ASSISTANT, AS QUOTED BY G. HOLTON (1971), P. 20

"But the years of anxious searching in the dark, with their intense longing, their alternations of confidence and exhaustion, and the final emergence into the light—only those who have experienced it can understand that"

EINSTEIN, AS QUOTED BY M. KLEIN (1971), P. 1315

"Of all the communities available to us there is not one I would want to devote myself to, except for the society of the true searchers, which has very few living members at any time. . ."

EINSTEIN LETTER TO BORN, QUOTED BY BORN (1971), P. 82

"I am studying your great works and—when I get stuck anywhere—now have the pleasure of seeing your friendly young face before me smiling and explaining"

EINSTEIN, LETTER OF MAY 2, 1920, AFTER MEETING NIELS BOHR

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."

EINSTEIN (1921), P. 28

"The most incomprehensible thing about the world is that it is comprehensible."

EINSTEIN, IN SCHILPP (1949), P. 112

EXERCISES

Exercise 1.1. CURVATURE OF A CYLINDER

Show that the Gaussian curvature R of the surface of a cylinder is zero by showing that geodesics on that surface (unroll!) suffer no geodesic deviation. Give an independent argument for the same conclusion by employing the formula $R = 1/\rho_1\rho_2$, where ρ_1 and ρ_2 are the principal radii of curvature at the point in question with respect to the enveloping Euclidean three-dimensional space.

Exercise 1.2. SPRING TIDE VS. NEAP TIDE

Evaluate (1) in conventional units and (2) in geometrized units the magnitude of the Newtonian tide-producing acceleration $R^m{}_{0n0}(m, n = 1, 2, 3)$ generated at the Earth by (1) the moon ($m_{\text{conv}} = 7.35 \times 10^{25}$ g, $r = 3.84 \times 10^{10}$ cm) and (2) the sun ($m_{\text{conv}} = 1.989 \times 10^{33}$ g, $r = 1.496 \times 10^{13}$ cm). By what factor do you expect spring tides to exceed neap tides?

Exercise 1.3. KEPLER ENCAPSULATED

A small satellite has a circular frequency $\omega(\text{cm}^{-1})$ in an orbit of radius r about a central object of mass $m(\text{cm})$. From the known value of ω, show that it is possible to determine neither r nor m individually, but only the effective "Kepler density" of the object as averaged over a sphere of the same radius as the orbit. Give the formula for ω^2 in terms of this Kepler density.

It is a reminder of the continuity of history that Kepler and Galileo (Box 1.9) wrote back and forth, and that the year that witnessed the death of Galileo saw the birth of Newton (Box 1.10). After Newton the first dramatically new synthesis of the laws of gravitation came from Einstein (Box 1.11).

And what the dead had no speech for, when living,
They can tell you, being dead; the communication
Of the dead is tongued with fire beyond
the language of the living.

T. S. ELIOT, in *LITTLE GIDDING* (1942)

I measured the skies
Now the shadows I measure
Skybound was the mind
Earthbound the body rests

JOHANNES KEPLER, d. November 15, 1630.
He wrote his epitaph in Latin;
it is translated by Coleman (1967), p. 109.

Ubi materia, ibi geometria.

JOHANNES KEPLER

PART II

PHYSICS IN FLAT SPACETIME

Wherein the reader meets an old friend, Special Relativity, outfitted in new, mod attire, and becomes more intimately acquainted with her charms.

CHAPTER 2

FOUNDATIONS OF SPECIAL RELATIVITY

In geometric and physical applications, it always turns out that a quantity is characterized not only by its tensor order, but also by symmetry.

HERMAN WEYL (1925)

Undoubtedly the most striking development of geometry during the last 2,000 years is the continual expansion of the concept "geometric object." This concept began by comprising only the few curves and surfaces of Greek synthetic geometry; it was stretched, during the Renaissance, to cover the whole domain of those objects defined by analytic geometry; more recently, it has been extended to cover the boundless universe treated by point-set theory.

KARL MENGER, IN SCHILPP (1949), P. 466.

§2.1. OVERVIEW

Curvature in geometry manifests itself as gravitation. Gravitation works on the separation of nearby particle world lines. In turn, particles and other sources of mass-energy cause curvature in the geometry. How does one break into this closed loop of the action of geometry on matter and the reaction of matter on geometry? One can begin no better than by analyzing the motion of particles and the dynamics of fields in a region of spacetime so limited that it can be regarded as flat. (See "Test for Flatness," Box 1.5).

Background assumed of reader

Chapters 2–6 develop this flat-spacetime viewpoint (special relativity). The reader, it is assumed, is already somewhat familiar with special relativity:* 4-vectors in general; the energy-momentum 4-vector; elementary Lorentz transformations; the Lorentz law for the force on a charged particle; at least one look at one equation

*For example, see Goldstein (1959), Leighton (1959), Jackson (1962), or, for the physical perspective presented geometrically, Taylor and Wheeler (1966).

in one book that refers to the electromagnetic field tensor $F_{\mu\nu}$; and the qualitative features of spacetime diagrams, including such points as (1) future and past light cones, (2) causal relationships ("past of," "future of," "neutral," or "in a spacelike relationship to"), (3) Lorentz contraction, (4) time dilation, (5) absence of a universal concept of simultaneity, and (6) the fact that the $\bar{t}$ and $\bar{z}$ axes in Box 2.4 are orthogonal even though they do not look so. If the reader finds anything new in these chapters, it will be: (i) a new viewpoint on special relativity, one emphasizing coordinate-free concepts and notation that generalize readily to curved spacetime ("geometric objects," tensors viewed as machines—treated in Chapters 2–4); or (ii) unfamiliar topics in special relativity, topics crucial to the later exposition of gravitation theory ("stress-energy tensor and conservation laws," Chapter 5; "accelerated observers," Chapter 6).

§2.2. GEOMETRIC OBJECTS

Every physical quantity can be described by a geometric object

Everything that goes on in spacetime has its geometric description, and almost every one of these descriptions lends itself to ready generalization from flat spacetime to curved spacetime. The greatest of the differences between one geometric object and another is its scope: the individual object (vector) for the momentum of a certain particle at a certain phase in its history, as contrasted to the extended geometric object that describes an electromagnetic field defined throughout space and time ("antisymmetric second-rank tensor field" or, more briefly, "field of 2-forms"). The idea that every physical quantity must be describable by a geometric object, and that the laws of physics must all be expressible as geometric relationships between these geometric objects, had its intellectual beginnings in the Erlanger program of Felix Klein (1872), came closer to physics in Einstein's "principle of general covariance" and in the writings of Hermann Weyl (1925), seems to have first been formulated clearly by Veblen and Whitehead (1932), and today pervades relativity theory, both special and general.

All laws of physics can be expressed geometrically

A. Nijenhuis (1952) and S.-S. Chern (1960, 1966, 1971) have expounded the mathematical theory of geometric objects. But to understand or do research in geometrodynamics, one need not master this elegant and beautiful subject. One need only know that geometric objects in spacetime are entities that exist independently of coordinate systems or reference frames. A point in spacetime ("*event*") is a geometric object. The arrow linking two neighboring events ("*vector*") is a geometric object in flat spacetime, and its generalization, the "*tangent vector,*" is a geometric object even when spacetime is curved. The "*metric*" (machine for producing the squared length of any vector; see Box 1.3) is a geometric object. No coordinates are needed to define any of these concepts.

The next few sections will introduce several geometric objects, and show the roles they play as representatives of physical quantities in flat spacetime.

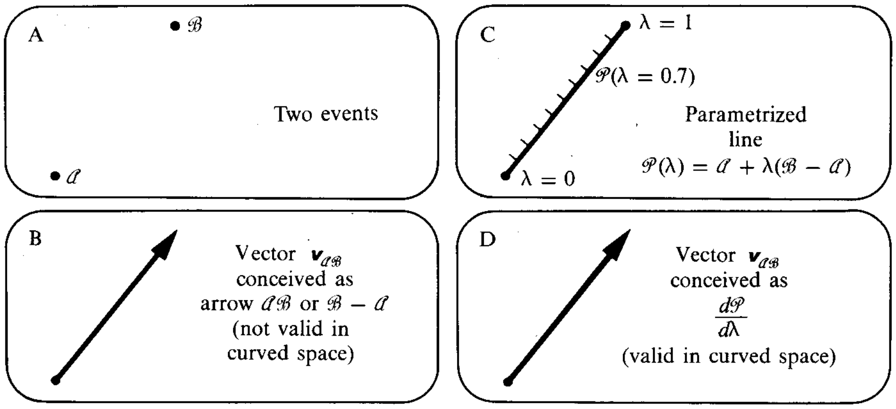


Figure 2.1.
From vector as connector of two points to vector as derivative ("tangent vector"; a local rather than a bilocal concept).

§2.3. VECTORS

Begin with the simplest idea of a vector (Figure 2.1B): an arrow extending from one spacetime event $\mathscr{A}$ ("tail") to another event $\mathscr{B}$ ("tip"). Write this vector as

Ways of defining vector:
As arrow

$$\boldsymbol{v}_{\mathscr{A}\mathscr{B}} = \mathscr{B} - \mathscr{A} \text{ (or } \mathscr{A}\mathscr{B}).$$

For many purposes (including later generalization to curved spacetime) other completely equivalent ways to think of this vector are more convenient. Represent the arrow by the parametrized straight line $\mathscr{P}(\lambda) = \mathscr{A} + \lambda(\mathscr{B} - \mathscr{A})$, with $\lambda = 0$ the tail of the arrow, and $\lambda = 1$ its tip. Form the derivative of this simple linear expression for $\mathscr{P}(\lambda)$:

As parametrized straight line

$$(d/d\lambda)[\mathscr{A} + \lambda(\mathscr{B} - \mathscr{A})] = \mathscr{B} - \mathscr{A} = \mathscr{P}(1) - \mathscr{P}(0) \equiv (\text{tip}) - (\text{tail}) \equiv \boldsymbol{v}_{\mathscr{A}\mathscr{B}}.$$

This result allows one to replace the idea of a vector as a 2-point object ("bilocal") by the concept of a vector as a 1-point object ("tangent vector"; local):

$$\boldsymbol{v}_{\mathscr{A}\mathscr{B}} = (d\mathscr{P}/d\lambda)_{\lambda=0}. \tag{2.1}$$

As derivative of point along curve

Example: if $\mathscr{P}(\tau)$ is the straight world line of a free particle, parametrized by its proper time, then the displacement that occurs in a proper time interval of one second gives an arrow $\boldsymbol{u} = \mathscr{P}(1) - \mathscr{P}(0)$. This arrow is easily drawn on a spacetime diagram. It accurately shows the 4-velocity of the particle. However, the derivative formula $\boldsymbol{u} = d\mathscr{P}/d\tau$ for computing the same displacement (1) is more suggestive of the velocity concept and (2) lends itself to the case of accelerated motion. Thus, given a world line $\mathscr{P}(\tau)$ that is not straight, as in Figure 2.2, one must first form $d\mathscr{P}/d\tau$, and only thereafter draw the straight line $\mathscr{P}(0) + \lambda(d\mathscr{P}/d\tau)_0$ of the arrow $\boldsymbol{u} = d\mathscr{P}/d\tau$ to display the 4-velocity $\boldsymbol{u}$.

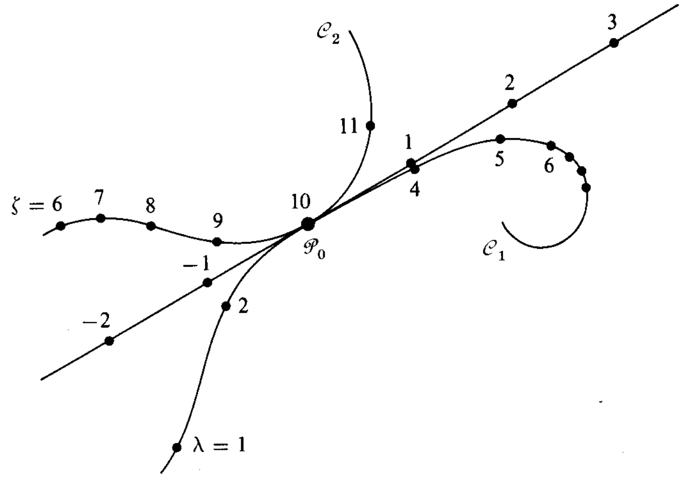


Figure 2.2.
Same tangent vector derived from two very different curves. That parametrized straight line is also drawn which best fits the two curves at $\mathcal{P}_0$. The tangent vector reaches from 0 to 1 on this straight line.

Components of a vector

The reader may be unfamiliar with this viewpoint. More familiar may be the components of the 4-velocity in a specific Lorentz reference frame:

$$u^0 = \frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \mathbf{v}^2}}, \qquad u^j = \frac{dx^j}{d\tau} = \frac{v^j}{\sqrt{1 - \mathbf{v}^2}}, \tag{2.2}$$

where

$$v^j = dx^j/dt = \text{components of "ordinary velocity,"}$$
$$\mathbf{v}^2 = (v^x)^2 + (v^y)^2 + (v^z)^2.$$

Even the components (2.2) of 4-velocity may seem slightly unfamiliar if the reader is accustomed to having the fourth component of a vector be multiplied by a factor $i = \sqrt{-1}$. If so, he must adjust himself to new notation. (See "Farewell to '*ict*,'" Box 2.1.)

Basis vectors

More fundamental than the components of a vector is the vector itself. It is a geometric object with a meaning independent of all coordinates. Thus a particle has a world line $\mathcal{P}(\tau)$, and a 4-velocity $\boldsymbol{u} = d\mathcal{P}/d\tau$, that have nothing to do with any coordinates. Coordinates enter the picture when analysis on a computer is required (rejects vectors; accepts numbers). For this purpose one adopts a Lorentz frame with orthonormal basis vectors (Figure 2.3) $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$. Relative to the origin $\mathcal{O}$ of this frame, the world line has a coordinate description

$$\mathcal{P}(\tau) - \mathcal{O} = x^0(\tau)\boldsymbol{e}_0 + x^1(\tau)\boldsymbol{e}_1 + x^2(\tau)\boldsymbol{e}_2 + x^3(\tau)\boldsymbol{e}_3 = x^\mu(\tau)\boldsymbol{e}_\mu.$$

Expressed relative to the same Lorentz frame, the 4-velocity of the particle is

$$\boldsymbol{u} = d\mathcal{P}/d\tau = (dx^\mu/d\tau)\boldsymbol{e}_\mu = u^0\boldsymbol{e}_0 + u^1\boldsymbol{e}_1 + u^2\boldsymbol{e}_2 + u^3\boldsymbol{e}_3. \tag{2.3}$$

Box 2.1 FAREWELL TO *"ict"*

One sometime participant in special relativity will have to be put to the sword: "$x^4 = ict$." This imaginary coordinate was invented to make the geometry of spacetime look formally as little different as possible from the geometry of Euclidean space; to make a Lorentz transformation look on paper like a rotation; and to spare one the distinction that one otherwise is forced to make between quantities with upper indices (such as the components p^μ of the energy-momentum vector) and quantities with lower indices (such as the components p_μ of the energy-momentum 1-form). However, it is no kindness to be spared this latter distinction. Without it, one cannot know whether a vector (§2.3) is meant or the very different geometric object that is a 1-form (§2.5). Moreover, there is a significant difference between an angle on which everything depends periodically (a rotation) and a parameter the increase of which gives rise to ever-growing momentum differences (the "velocity parameter" of a Lorentz transformation; Box 2.4). If the imaginary time-coordinate hides from view the character of the geometric object being dealt with and the nature of the parameter in a transformation, it also does something even more serious: it hides the completely different metric structure (§2.4) of $+++$ geometry and $-+++$ geometry. In Euclidean geometry, when the distance between two points is zero, the two points must be the same point. In Lorentz-Minkowski geometry, when the interval between two events is zero, one event may be on Earth and the other on a supernova in the galaxy M31, but their separation must be a null ray (piece of a light cone). The backward-pointing light cone at a given event contains all the events by which that event can be influenced. The forward-pointing light cone contains all events that it can influence. The multitude of double light cones taking off from all the events of spacetime forms an interlocking causal structure. This structure makes the machinery of the physical world function as it does (further comments on this structure in Wheeler and Feynman 1945 and 1949 and in Zeeman 1964). If in a region where spacetime is flat, one can hide this structure from view by writing

$$(\Delta s)^2 = (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 + (\Delta x^4)^2,$$

with $x^4 = ict$, no one has discovered a way to make an imaginary coordinate work in the general curved spacetime manifold. If "$x^4 = ict$" cannot be used there, it will not be used here. In this chapter and hereafter, as throughout the literature of general relativity, a real time coordinate is used, $x^0 = t = ct_{\text{conv}}$ (superscript 0 rather than 4 to avoid any possibility of confusion with the imaginary time coordinate).

The components w^α of any other vector $\boldsymbol{w}$ in this frame are similarly defined as the coefficients in such an expansion,

Expansion of vector in terms of basis

$$\boldsymbol{w} = w^\alpha \boldsymbol{e}_\alpha. \tag{2.4}$$

Notice: the subscript α on $\boldsymbol{e}_\alpha$ tells which vector, not which component!

§2.4. THE METRIC TENSOR

The metric tensor, one recalls from part IV of Box 1.3, is a machine for calculating the squared length of a single vector, or the scalar product of two different vectors.

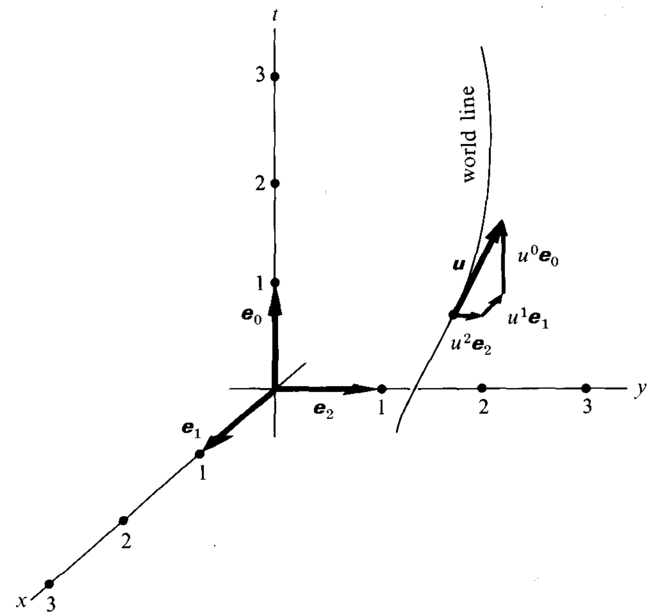


Figure 2.3.
The 4-velocity of a particle in flat spacetime. The 4-velocity $\boldsymbol{u}$ is the unit vector (arrow) tangent to the particle's world line—one tangent vector for each event on the world line. In a specific Lorentz coordinate system, there are basis vectors of unit length, which point along the four coordinate axes: $\boldsymbol{e}_0$, $\boldsymbol{e}_1$,$\boldsymbol{e}_2$, $\boldsymbol{e}_3$. The 4-velocity, like any vector, can be expressed as a sum of components along the basis vectors:

$$\boldsymbol{u} = u^0\boldsymbol{e}_0 + u^1\boldsymbol{e}_1 + u^2\boldsymbol{e}_2 + u^3\boldsymbol{e}_3 = u^\alpha \boldsymbol{e}_\alpha.$$

Metric defined as machine for computing scalar products of vectors

More precisely, the metric tensor $\boldsymbol{g}$ is a machine with two slots for inserting vectors

$$\begin{array}{c} \text{slot 1} \quad \text{slot 2} \\ \downarrow \qquad \downarrow \\ \boldsymbol{g}(\qquad , \qquad). \end{array} \tag{2.5}$$

Upon insertion, the machine spews out a real number:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) &= \text{"scalar product of } \boldsymbol{u} \text{ and } \boldsymbol{v}\text{," also denoted } \boldsymbol{u}\cdot\boldsymbol{v}. \\ \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) &= \text{"squared length of } \boldsymbol{u}\text{," also denoted } \boldsymbol{u}^2. \end{aligned} \tag{2.6}$$

Moreover, this number is independent of the order in which the vectors are inserted ("symmetry of metric tensor"),

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{u}); \tag{2.7}$$

and it is linear in the vectors inserted

$$\boldsymbol{g}(a\boldsymbol{u} + b\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{g}(\boldsymbol{w}, a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{w}) + b\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}). \tag{2.8}$$

Because the metric "machine" is linear, one can calculate its output, for any input, as follows, if one knows only what it does to the basis vectors $\boldsymbol{e}_\alpha$ of a Lorentz frame. (1) Define the symbols ("metric coefficients") $\eta_{\alpha\beta}$ by

Metric coefficients

$$\eta_{\alpha\beta} \equiv \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) = \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta. \tag{2.9}$$

(2) Calculate their numerical values from the known squared length of the separation vector $\boldsymbol{\xi} = \Delta x^\alpha \boldsymbol{e}_\alpha$ between two events:

$$\begin{aligned}(\Delta s)^2 &= -(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 \\ &= \boldsymbol{g}(\Delta x^\alpha \boldsymbol{e}_\alpha, \Delta x^\beta \boldsymbol{e}_\beta) = \Delta x^\alpha \Delta x^\beta \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) \\ &= \Delta x^\alpha \Delta x^\beta \eta_{\alpha\beta} \qquad \text{for every choice of } \Delta x^\alpha\end{aligned}$$

$$\Longrightarrow \|\eta_{\alpha\beta}\| \equiv \begin{Vmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix} \text{ in any Lorentz frame.} \tag{2.10}$$

(3) Calculate the scalar product of any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ from

$$\begin{aligned}\boldsymbol{u} \cdot \boldsymbol{v} &= \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(u^\alpha \boldsymbol{e}_\alpha, v^\beta \boldsymbol{e}_\beta) = u^\alpha v^\beta \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta); \\ \boldsymbol{u} \cdot \boldsymbol{v} &= u^\alpha v^\beta \eta_{\alpha\beta} = -u^0 v^0 + u^1 v^1 + u^2 v^2 + u^3 v^3.\end{aligned} \tag{2.11}$$

Scalar products computed from components of vectors

That one can classify directions and vectors in spacetime into *"timelike"* (negative squared length), *"spacelike"* (positive squared length), and *"null"* or *"lightlike"* (zero squared length) is made possible by the negative sign on the metric coefficient η_{00}.

Box 2.2 shows applications of the above ideas and notation to two elementary problems in special relativity theory.

§2.5. DIFFERENTIAL FORMS

Vectors and the metric tensor are geometric objects that are already familiar from Chapter 1 and from elementary courses in special relativity. Not so familiar, yet equally important, is a third geometric object: the *"differential form"* or *"1-form."*

Consider the 4-momentum $\boldsymbol{p}$ of a particle, an electron, for example. To spell out one concept of momentum, start with the 4-velocity, $\boldsymbol{u} = d\mathscr{P}/d\tau$, of this electron ("spacetime displacement per unit of proper time along a straightline approximation of the world line"). This is a vector of unit length. Multiply by the mass m of the particle to obtain the *momentum vector*

$$\boldsymbol{p} = m\boldsymbol{u}.$$

The 1-form illustrated by de Broglie waves

But physics gives also quite another idea of momentum. It associates a de Broglie wave with each particle. Moreover, this wave has the most direct possible physical significance. Diffract this wave from a crystal lattice. From the pattern of diffraction, one can determine not merely the length of the de Broglie waves, but also the pattern in space made by surfaces of equal, integral phase $\phi = 7$, $\phi = 8$, $\phi = 9$, ... This

Box 2.2 WORKED EXERCISES USING THE METRIC

Exercise: Show that the squared length of a test particle's 4-velocity $\boldsymbol{u}$ is -1. *Solution:* In any Lorentz frame, using the components (2.2), one calculates as follows

$$\begin{aligned}\boldsymbol{u}^2 = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = u^\alpha u^\beta \eta_{\alpha\beta} &= -(u^0)^2 + (u^1)^2 + (u^2)^2 + (u^3)^2 \\ &= -\frac{1}{1-v^2} + \frac{v^2}{1-v^2} = -1.\end{aligned}$$

Exercise: Show that the rest mass of a particle is related to its energy and momentum by the famous equation

$$(mc^2)^2 = E^2 - (\boldsymbol{p}c)^2$$

or, equivalently (geometrized units!),

$$m^2 = E^2 - \boldsymbol{p}^2.$$

First Solution: The 4-momentum is defined by $\boldsymbol{p} = m\boldsymbol{u}$, where $\boldsymbol{u}$ is the 4-velocity and m is the rest mass. Consequently, its squared length is

$$\begin{aligned}\boldsymbol{p}^2 &= m^2\boldsymbol{u}^2 = -m^2 \\ &= -(mu^0)^2 + m^2\boldsymbol{u}^2 = \underset{\substack{\uparrow \\ E^2}}{-\frac{m^2}{1-v^2}} + \underset{\substack{\uparrow \\ \boldsymbol{p}^2}}{\frac{m^2v^2}{1-v^2}}.\end{aligned}$$

Second Solution: In the frame of the observer, where E and $\boldsymbol{p}$ are measured, the 4-momentum splits into time and space parts as

$$p^0 = E, \qquad p^1\boldsymbol{e}_1 + p^2\boldsymbol{e}_2 + p^3\boldsymbol{e}_3 = \boldsymbol{p};$$

hence, its squared length is

$$\boldsymbol{p}^2 = -E^2 + \boldsymbol{p}^2.$$

But in the particle's rest frame, $\boldsymbol{p}$ splits as

$$p^0 = m, \qquad p^1 = p^2 = p^3 = 0;$$

hence, its squared length is $\boldsymbol{p}^2 = -m^2$. But the squared length is a geometric object defined independently of any coordinate system; so it must be the same by whatever means one calculates it:

$$-\boldsymbol{p}^2 = m^2 = E^2 - \boldsymbol{p}^2.$$

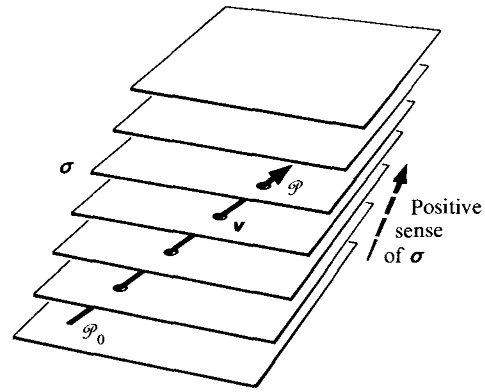


Figure 2.4.
The vector separation $\boldsymbol{v} = \mathscr{P} - \mathscr{P}_0$ between two neighboring events $\mathscr{P}_0$ and $\mathscr{P}$; a 1-form $\boldsymbol{\sigma}$; and the piercing of $\boldsymbol{\sigma}$ by $\boldsymbol{v}$ to give the number

$$\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle = (\text{number of surfaces pierced}) = 4.4$$

(4.4 "bongs of bell"). When $\boldsymbol{\sigma}$ is made of surfaces of constant phase, $\phi = 17$, $\phi = 18$, $\phi = 19, \ldots$ of the de Broglie wave for an electron, then $\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle$ is the phase difference between the events $\mathscr{P}_0$ and $\mathscr{P}$. Note that $\boldsymbol{\sigma}$ is not fully specified by its surfaces; an orientation is also necessary. Which direction from surface to surface is "positive"; i.e., in which direction does ϕ increase?

pattern of surfaces, given a name "$\tilde{\boldsymbol{k}}$," provides the simplest illustration one can easily find for a 1-form.

Vector pierces 1-form

The pattern of surfaces in spacetime made by such a 1-form: what is it good for? Take two nearby points in spacetime, $\mathscr{P}$ and $\mathscr{P}_0$. Run an arrow $\boldsymbol{v} = \mathscr{P} - \mathscr{P}_0$ from $\mathscr{P}_0$ to $\mathscr{P}$. It will pierce a certain number of the de Broglie wave's surfaces of integral phase, with a bong of an imaginary bell at each piercing. The number of surfaces pierced (number of "bongs of bell") is denoted

$$\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle;$$

1-form pierced ⟶ ⟵ vector that pierces

in this example it equals the phase difference between tail ($\mathscr{P}_0$) and tip ($\mathscr{P}$) of $\boldsymbol{v}$,

$$\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle = \phi(\mathscr{P}) - \phi(\mathscr{P}_0).$$

See Figure 2.4.

Normally neither $\mathscr{P}_0$ nor $\mathscr{P}$ will lie at a point of integral phase. Therefore one can and will imagine, as uniformly interpolated between the surfaces of integral phase, an infinitude of surfaces with all the intermediate phase values. With their aid, the precise value of $\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle = \phi(\mathscr{P}) - \phi(\mathscr{P}_0)$ can be determined.

The 1-form viewed as family of flat, equally spaced surfaces

To make the mathematics simple, regard $\tilde{\boldsymbol{k}}$ not as the global pattern of de Broglie-wave surfaces, but as a local pattern near a specific point in spacetime. Just as the vector $\boldsymbol{u} = d\mathscr{P}/d\tau$ represents the local behavior of a particle's world line (linear approximation to curved line in general), so the 1-form $\tilde{\boldsymbol{k}}$ represents the local form

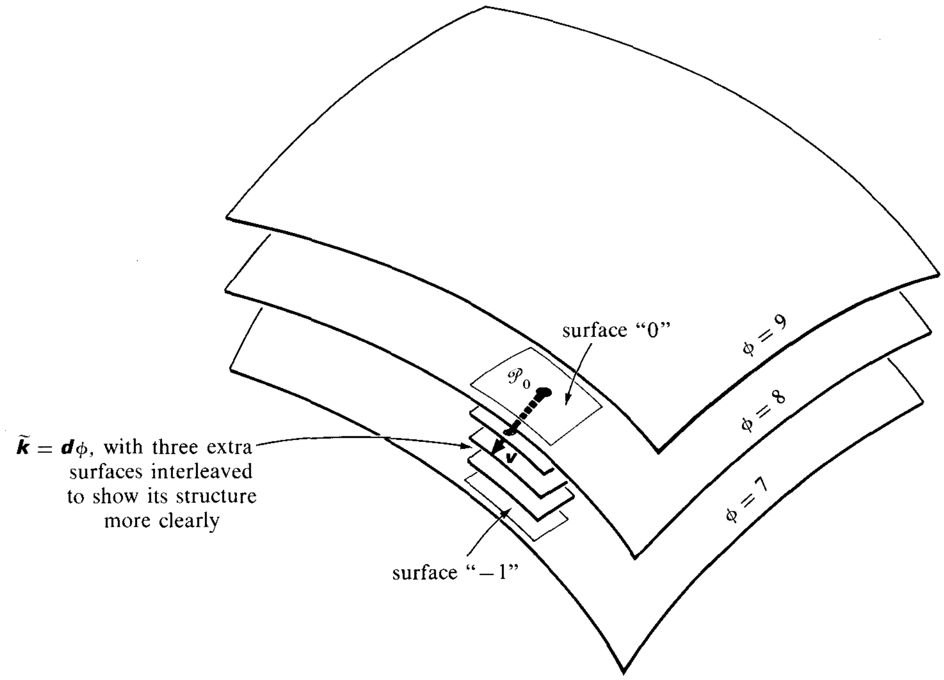


Figure 2.5.
This is a dual-purpose figure. (a) It illustrates the de Broglie wave 1-form $\tilde{\boldsymbol{k}}$ at an event $\mathscr{P}_0$ (family of equally spaced, flat surfaces, or "hyperplanes" approximating the surfaces of constant phase). (b) It illustrates the gradient $\boldsymbol{d}\phi$ of the function ϕ (concept defined in §2.6), which is the same oriented family of flat surfaces

$$\tilde{\boldsymbol{k}} = \boldsymbol{d}\phi.$$

At different events, $\tilde{\boldsymbol{k}} = \boldsymbol{d}\phi$ is different—different orientation of surfaces and different spacing. The change in ϕ between the tail and tip of the very short vector $\boldsymbol{v}$ is equal to the number of surfaces of $\boldsymbol{d}\phi$ pierced by $\boldsymbol{v}$, $\langle \boldsymbol{d}\phi, \boldsymbol{v}\rangle$; it equals -0.5 in this figure.

of the de Broglie wave's surfaces (linear approximation; surfaces flat and equally spaced; see Figure 2.5).

Regard the 1-form $\tilde{\boldsymbol{k}}$ as a machine into which vectors are inserted, and from which numbers emerge. Insertion of $\boldsymbol{v}$ produces as output $\langle \tilde{\boldsymbol{k}}, \boldsymbol{v}\rangle$. Since the surfaces of $\tilde{\boldsymbol{k}}$ are flat and equally spaced, the output is a linear function of the input:

$$\langle \tilde{\boldsymbol{k}}, a\boldsymbol{u} + b\boldsymbol{v}\rangle = a\langle \tilde{\boldsymbol{k}}, \boldsymbol{u}\rangle + b\langle \tilde{\boldsymbol{k}}, \boldsymbol{v}\rangle. \tag{2.12a}$$

The 1-form viewed as linear function of vectors

This, in fact, is the mathematical definition of a 1-form: *a 1-form is a linear, real-valued function of vectors;* i.e., a linear machine that takes in a vector and puts out a number. Given the machine $\tilde{\boldsymbol{k}}$, it is straightforward to draw the corresponding surfaces in spacetime. Pick a point $\mathscr{P}_0$ at which the machine is to reside. The surface of $\tilde{\boldsymbol{k}}$ that passes through $\mathscr{P}_0$ contains points $\mathscr{P}$ for which $\langle \tilde{\boldsymbol{k}}, \mathscr{P} - \mathscr{P}_0\rangle = 0$ (no bongs of bell). The other surfaces contain points with $\langle \tilde{\boldsymbol{k}}, \mathscr{P} - \mathscr{P}_0\rangle = \pm 1, \pm 2, \pm 3, \ldots$.

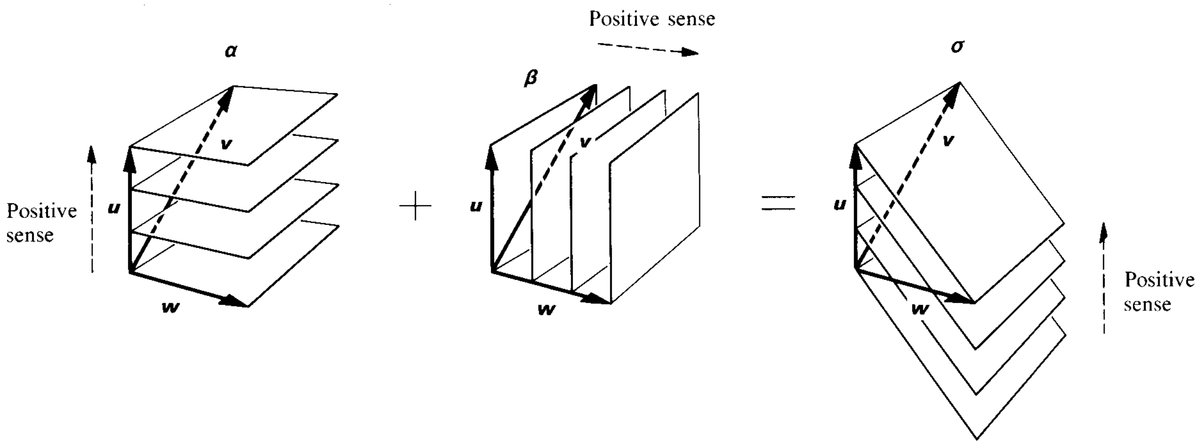


Figure 2.6.
The addition of two 1-forms, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, to produce the 1-form $\boldsymbol{\sigma}$. Required is a pictorial construction that starts from the surfaces of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, e.g., $\langle \boldsymbol{\alpha}, \mathscr{P} - \mathscr{P}_0 \rangle = \cdots -1, 0, 1, 2, \ldots$, and constructs those of $\boldsymbol{\sigma} = \boldsymbol{\alpha} + \boldsymbol{\beta}$. Such a construction, based on linearity (2.12b) of the addition process, is as follows. (1) Pick several vectors $\boldsymbol{u}, \boldsymbol{v}, \ldots$ that lie parallel to the surfaces of $\boldsymbol{\beta}$ (no piercing!), but pierce precisely 3 surfaces of $\boldsymbol{\alpha}$; each of these must then pierce precisely 3 surfaces of $\boldsymbol{\sigma}$:

$$\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle = \langle \boldsymbol{\alpha} + \boldsymbol{\beta}, \boldsymbol{u} \rangle = \langle \boldsymbol{\alpha}, \boldsymbol{u} \rangle = 3.$$

(2) Pick several other vectors $\boldsymbol{w}, \ldots$ that lie parallel to the surfaces of $\boldsymbol{\alpha}$ but pierce precisely 3 surfaces of $\boldsymbol{\beta}$; these will also pierce precisely 3 surfaces of $\boldsymbol{\sigma}$. (3) Construct that unique family of equally spaced surfaces in which $\boldsymbol{u}, \boldsymbol{v}, \ldots, \boldsymbol{w}, \ldots$ all have their tails on one surface and their tips on the third succeeding surface.

Sometimes 1-forms are denoted by boldface, sans-serif Latin letters with tildes over them, e.g., $\tilde{\mathbf{k}}$; but more often by boldface Greek letters, e.g., $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}$. The output of a 1-form $\boldsymbol{\sigma}$, when a vector $\boldsymbol{u}$ is inserted, is called *"the value of $\boldsymbol{\sigma}$ on $\boldsymbol{u}$"* or *"the contraction of $\boldsymbol{\sigma}$ with $\boldsymbol{u}$."*

Addition of 1-forms

Also, 1-forms, like any other kind of function, can be added. The 1-form $a\boldsymbol{\alpha} + b\boldsymbol{\beta}$ is that machine (family of surfaces) which puts out the following number when a vector $\boldsymbol{u}$ is put in:

$$\langle a\boldsymbol{\alpha} + b\boldsymbol{\beta}, \boldsymbol{u} \rangle = a\langle \boldsymbol{\alpha}, \boldsymbol{u} \rangle + b\langle \boldsymbol{\beta}, \boldsymbol{u} \rangle. \tag{2.12b}$$

Figure 2.6 depicts this addition in terms of surfaces.

One can verify that the set of all 1-forms at a given event is a "vector space" in the abstract, algebraic sense of the term.

Return to a particle and its de Broglie wave. Just as the arrow $\boldsymbol{p} = m d\mathscr{P}/d\tau$ represents the best *linear* approximation to the particle's actual world line near $\mathscr{P}_0$, so the flat surfaces of the 1-form $\tilde{\mathbf{k}}$ provide the best linear approximation to the curved surfaces of the particle's de Broglie wave, and $\tilde{\mathbf{k}}$ itself is the *linear function* that best approximates the de Broglie phase ϕ near $\mathscr{P}_0$:

$$\begin{aligned}\phi(\mathscr{P}) = \phi(\mathscr{P}_0) &+ \langle \tilde{\mathbf{k}}, \mathscr{P} - \mathscr{P}_0 \rangle \\ &+ \text{terms of higher order in } (\mathscr{P} - \mathscr{P}_0).\end{aligned} \tag{2.13}$$

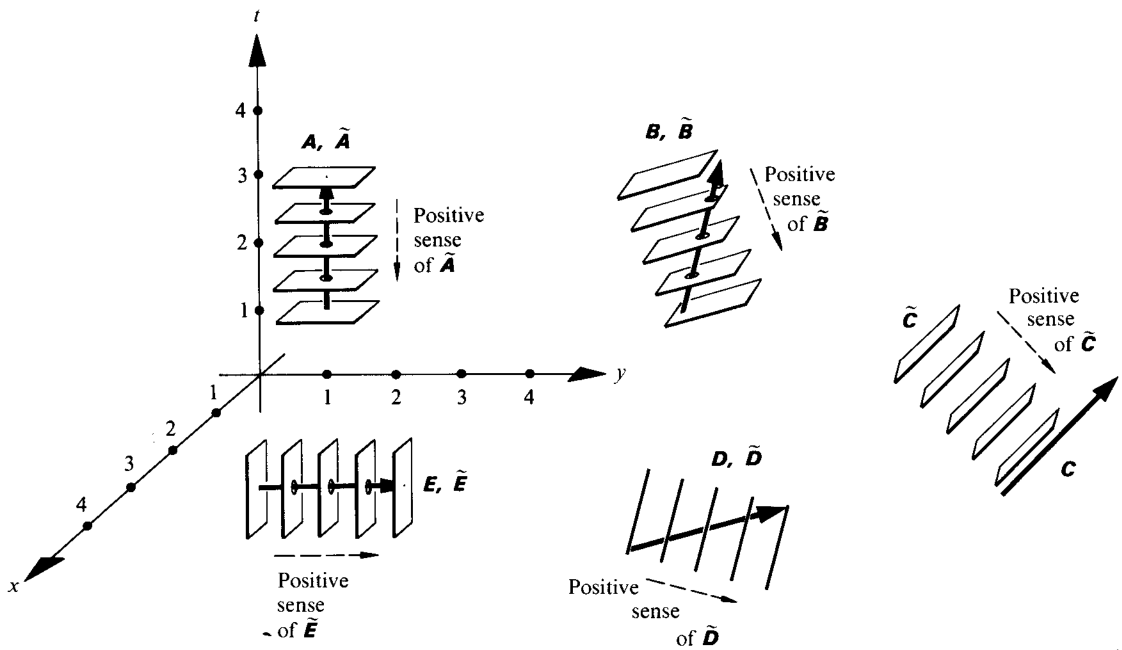


Figure 2.7.
Several vectors, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$, and corresponding 1-forms $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{C}}$, $\tilde{\boldsymbol{D}}$, $\tilde{\boldsymbol{E}}$. The process of drawing $\tilde{\boldsymbol{U}}$ corresponding to a given vector $\boldsymbol{U}$ is quite simple. (1) Orient the surfaces of $\tilde{\boldsymbol{U}}$ orthogonal to the vector $\boldsymbol{U}$. (Why? Because any vector $\boldsymbol{V}$ that is perpendicular to $\boldsymbol{U}$ must pierce no surfaces of $\tilde{\boldsymbol{U}}$ ($0 = \boldsymbol{U} \cdot \boldsymbol{V} = \langle \tilde{\boldsymbol{U}}, \boldsymbol{V} \rangle$) and must therefore lie in a surface of $\tilde{\boldsymbol{U}}$.) (2) Space the surfaces of $\tilde{\boldsymbol{U}}$ so that the number of surfaces pierced by some arbitrary vector $\boldsymbol{Y}$ (e.g., $\boldsymbol{Y} = \boldsymbol{U}$) is equal to $\boldsymbol{Y} \cdot \boldsymbol{U}$.

Note that in the figure the surfaces of $\tilde{\boldsymbol{B}}$ are, indeed, orthogonal to $\boldsymbol{B}$; those of $\tilde{\boldsymbol{C}}$ are, indeed, orthogonal to $\boldsymbol{C}$, etc. If they do not look so, that is because the reader is attributing Euclidean geometry, not Lorentz geometry, to the spacetime diagram. He should recall, for example, that because $\boldsymbol{C}$ is a null vector, it is orthogonal to itself ($\boldsymbol{C} \cdot \boldsymbol{C} = 0$), so it must itself lie in a surface of the 1-form $\tilde{\boldsymbol{C}}$. Confused readers may review spacetime diagrams in a more elementary text, e.g., Taylor and Wheeler (1966).

Physical correspondence between 1-forms and vectors

Actually, the de Broglie 1-form $\tilde{\boldsymbol{k}}$ and the momentum vector $\boldsymbol{p}$ contain precisely the same information, both physically (via quantum theory) and mathematically. To see their relationship. relabel the surfaces of $\tilde{\boldsymbol{k}}$ by $\hbar \times$ phase, thereby obtaining the *"momentum 1-form"* $\tilde{\boldsymbol{p}}$. Pierce this 1-form with any vector $\boldsymbol{v}$, and find the result (exercise 2.1) that

$$\boldsymbol{p} \cdot \boldsymbol{v} = \langle \tilde{\boldsymbol{p}}, \boldsymbol{v} \rangle. \tag{2.14}$$

In words: the projection of $\boldsymbol{v}$ on the 4-momentum vector $\boldsymbol{p}$ equals the number of surfaces it pierces in the 4-momentum 1-form $\tilde{\boldsymbol{p}}$. *Examples:* Vectors $\boldsymbol{v}$ lying in a surface of $\tilde{\boldsymbol{p}}$ (no piercing) are perpendicular to $\boldsymbol{p}$ (no projection); $\boldsymbol{p}$ itself pierces $\boldsymbol{p}^2 = -m^2$ surfaces of $\tilde{\boldsymbol{p}}$.

Mathematical correspondence between 1-forms and vectors

Corresponding to any vector $\boldsymbol{p}$ there exists a unique 1-form (linear function of vectors) $\tilde{\boldsymbol{p}}$ defined by equation (2.14). And corresponding to any 1-form $\tilde{\boldsymbol{p}}$, there exists a unique vector $\boldsymbol{p}$ defined by its projections on all other vectors, by equation (2.14). Figure 2.7 shows several vectors and their corresponding 1-forms.

A single physical quantity can be described equally well by a vector $\boldsymbol{p}$ or by the corresponding 1-form $\tilde{\boldsymbol{p}}$. Sometimes the vector description is the simplest and most natural; sometimes the 1-form description is nicer. *Example:* Consider a 1-form representing the march of Lorentz coordinate time toward the future—surfaces $x^0 = \ldots, 7, 8, 9, \ldots$. The corresponding vector points toward the past [see Figure 2.7 or equation (2.14)]; its description of the forward march of time is not so nice!

One often omits the tilde from the 1-form $\tilde{\boldsymbol{p}}$ corresponding to a vector $\boldsymbol{p}$, and uses the same symbol $\boldsymbol{p}$ for both. Such practice is justified by the unique correspondence (both mathematical and physical) between $\tilde{\boldsymbol{p}}$ and $\boldsymbol{p}$.

EXERCISE

Exercise 2.1.
Show that equation (2.14) is in accord with the quantum-mechanical properties of a de Broglie wave,

$$\psi = e^{i\phi} = \exp\left[i(\boldsymbol{k} \cdot \boldsymbol{x} - \omega t)\right].$$

---

§2.6. GRADIENTS AND DIRECTIONAL DERIVATIVES

Gradient of a function as a 1-form

There is no simpler 1-form than the *gradient*, "$\boldsymbol{d}f$," of a function f. Gradient a 1-form? How so? Hasn't one always known the gradient as a vector? Yes, indeed, but only because one was not familiar with the more appropriate 1-form concept. The more familiar gradient is the vector corresponding, via equation (2.14), to the 1-form gradient. The hyperplanes representing $\boldsymbol{d}f$ at a point $\mathscr{P}_0$ are just the level surfaces of f itself, except for flattening and adjustment to equal spacing (Figure 2.5; identify f here with ϕ there). More precisely, they are the level surfaces of the linear function that approximates f in an infinitesimal neighborhood of $\mathscr{P}_0$.

Why the name "gradient"? Because $\boldsymbol{d}f$ describes the first order changes in f in the neighborhood of $\mathscr{P}_0$:

$$f(\mathscr{P}) = f(\mathscr{P}_0) + \langle \boldsymbol{d}f, \mathscr{P} - \mathscr{P}_0 \rangle + (\text{nonlinear terms}). \tag{2.15}$$

[Compare the fundamental idea of "derivative" of something as "best linear approximation to that something at a point"—an idea that works even for functions whose values and arguments are infinite dimensional vectors! See, e.g., Dieudonné (1960).]

Take any vector $\boldsymbol{v}$; construct the curve $\mathscr{P}(\lambda)$ defined by $\mathscr{P}(\lambda) - \mathscr{P}_0 = \lambda \boldsymbol{v}$; and differentiate the function f along this curve:

$$\partial_{\boldsymbol{v}} f = (d/d\lambda)_{\lambda=0} f[\mathscr{P}(\lambda)] = (df/d\lambda)_{\mathscr{P}_0}. \tag{2.16a}$$

The "differential operator,"

$$\partial_{\boldsymbol{v}} = (d/d\lambda)_{\text{at } \lambda = 0, \text{ along curve } \mathscr{P}(\lambda) - \mathscr{P}_0 = \lambda \boldsymbol{v}}, \tag{2.16b}$$

Directional derivative operator defined

which does this differentiating, is called the *"directional derivative operator along the vector* $\boldsymbol{v}$." The directional derivative $\partial_{\boldsymbol{v}} f$ and the gradient $\boldsymbol{d}f$ are intimately related, as one sees by applying $\partial_{\boldsymbol{v}}$ to equation (2.15) and evaluating the result at the point $\mathscr{P}_0$:

$$\partial_{\boldsymbol{v}} f = \langle \boldsymbol{d}f, d\mathscr{P}/d\lambda \rangle = \langle \boldsymbol{d}f, \boldsymbol{v} \rangle. \tag{2.17}$$

This result, expressed in words, is: $\boldsymbol{d}f$ is a linear machine for computing the rate of change of f along any desired vector $\boldsymbol{v}$. Insert $\boldsymbol{v}$ into $\boldsymbol{d}f$; the output ("number of surfaces pierced; number of bongs of bell") is $\partial_{\boldsymbol{v}} f$—which, for sufficiently small $\boldsymbol{v}$, is simply the difference in f between tip and tail of $\boldsymbol{v}$.

§2.7. COORDINATE REPRESENTATION OF GEOMETRIC OBJECTS

In flat spacetime, special attention focuses on Lorentz frames. The coordinates $x^0(\mathscr{P})$, $x^1(\mathscr{P})$, $x^2(\mathscr{P})$, $x^3(\mathscr{P})$ of a Lorentz frame are functions; so their gradients can be calculated. Each of the resulting "basis 1-forms,"

Basis 1-forms

$$\boldsymbol{\omega}^\alpha = \boldsymbol{d}x^\alpha, \tag{2.18}$$

has as its hyperplanes the coordinate surfaces $x^\alpha = \text{const}$; see Figure 2.8. Consequently the basis vector $\boldsymbol{e}_\alpha$ pierces precisely one surface of the basis 1-form $\boldsymbol{\omega}^\alpha$,

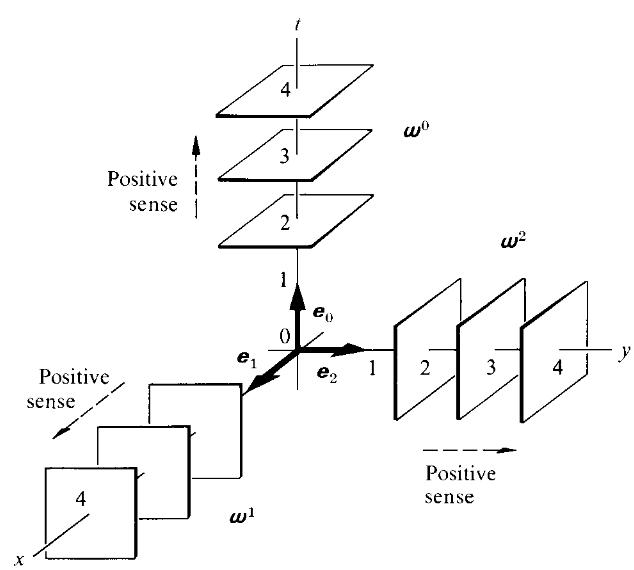


Figure 2.8.
The basis vectors and 1-forms of a particular Lorentz coordinate frame. The basis 1-forms are so laid out that

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta.$$

while the other three basis vectors lie parallel to the surfaces of $\boldsymbol{\omega}^\alpha$ and thus pierce none:

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta. \tag{2.19}$$

(One says that the set of basis 1-forms $\{\boldsymbol{\omega}^\alpha\}$ and the set of basis vectors $\{\boldsymbol{e}_\beta\}$ are the *"duals"* of each other if they have this property.)

Just as arbitrary vectors can be expanded in terms of the basis $\boldsymbol{e}_\alpha$, $\boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha$, so arbitrary 1-forms can be expanded in terms of $\boldsymbol{\omega}^\beta$:

Expansion of 1-form in terms of basis

$$\boldsymbol{\sigma} = \sigma_\beta \boldsymbol{\omega}^\beta. \tag{2.20}$$

The expansion coefficients σ_β are called "the components of $\boldsymbol{\sigma}$ on the basis $\boldsymbol{\omega}^\beta$."

These definitions produce an elegant computational formalism, thus: Calculate how many surfaces of $\boldsymbol{\sigma}$ are pierced by the basis vector $\boldsymbol{e}_\alpha$; equations (2.19) and (2.20) give the answer:

Calculation and manipulation of vector and 1-form components

$$\langle \boldsymbol{\sigma}, \boldsymbol{e}_\alpha \rangle = \langle \sigma_\beta \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \sigma_\beta \langle \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \sigma_\beta \delta^\beta{}_\alpha;$$

i.e.,

$$\langle \boldsymbol{\sigma}, \boldsymbol{e}_\alpha \rangle = \sigma_\alpha. \tag{2.21a}$$

Similarly, calculate $\langle \boldsymbol{\omega}^\alpha, \boldsymbol{v} \rangle$ for any vector $\boldsymbol{v} = \boldsymbol{e}_\beta v^\beta$; the result is

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{v} \rangle = v^\alpha. \tag{2.21b}$$

Multiply equation (2.21a) by v^α and sum, or multiply (2.21b) by σ_α and sum; the result in either case is

$$\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle = \sigma_\alpha v^\alpha. \tag{2.22}$$

This provides a way, using components, to calculate the coordinate-independent value of $\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle$.

Each Lorentz frame gives a coordinate-dependent representation of any geometric object or relation: $\boldsymbol{v}$ is represented by its components v^α; $\boldsymbol{\sigma}$, by its components σ_α; a point $\mathscr{P}$, by its coordinates x^α; the relation $\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle = 17.3$ by $\sigma_\alpha v^\alpha = 17.3$.

To find the coordinate representation of the directional derivative operator $\partial_{\boldsymbol{v}}$, rewrite equation (2.16b) using elementary calculus

$$\partial_{\boldsymbol{v}} = \left(\frac{d}{d\lambda}\right)_{\mathscr{P}_0} = \underbrace{\left(\frac{dx^\alpha}{d\lambda}\right)_{\text{at } \mathscr{P}_0 \text{ along } \mathscr{P}(\lambda) - \mathscr{P}_0 = \lambda \boldsymbol{v}}}_{v^\alpha;\ \text{see equation (2.3)}} \left(\frac{\partial}{\partial x^\alpha}\right);$$

the result is

$$\partial_{\boldsymbol{v}} = v^\alpha\, \partial/\partial x^\alpha. \tag{2.23}$$

Directional derivative in terms of coordinates

In particular, the directional derivative along a basis vector $\boldsymbol{e}_\alpha$ (components $[\boldsymbol{e}_\alpha]^\beta = \langle \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \delta^\beta{}_\alpha$) is

$$\partial_\alpha \equiv \partial_{\boldsymbol{e}_\alpha} = \partial/\partial x^\alpha. \tag{2.24}$$

This should also be obvious from Figure 2.8.

Components of gradient

The components of the gradient 1-form $\boldsymbol{d}f$, which are denoted $f_{,\alpha}$

$$\boldsymbol{d}f = f_{,\alpha}\boldsymbol{\omega}^{\alpha}, \tag{2.25a}$$

are calculated easily using the above formulas:

$$\begin{aligned} f_{,\alpha} &= \langle \boldsymbol{d}f, \boldsymbol{e}_{\alpha}\rangle \text{ [standard way to calculate components; equation (2.21a)]} \\ &= \partial_{\alpha} f \quad \text{[by relation (2.17) between directional derivative and gradient]} \\ &= \partial f/\partial x^{\alpha} \quad \text{[by equation (2.24)].} \end{aligned}$$

Thus, in agreement with the elementary calculus idea of gradient, the components of $\boldsymbol{d}f$ are just the partial derivatives along the coordinate axes:

$$f_{,\alpha} = \partial f/\partial x^{\alpha}; \qquad \text{i.e., } \boldsymbol{d}f = (\partial f/\partial x^{\alpha})\, \boldsymbol{d}x^{\alpha}. \tag{2.25b}$$

(Recall: $\boldsymbol{\omega}^{\alpha} = \boldsymbol{d}x^{\alpha}$.) The formula $\boldsymbol{d}f = (\partial f/\partial x^{\alpha})\, \boldsymbol{d}x^{\alpha}$ suggests, correctly, that $\boldsymbol{d}f$ is a rigorous version of the "differential" of elementary calculus; see Box 2.3.

Other important coordinate representations for geometric relations are explored in the following exercises.

EXERCISES

Derive the following computationally useful formulas:

Exercise 2.2. LOWERING INDEX TO GET THE 1-FORM CORRESPONDING TO A VECTOR

The components u_{α} of the 1-form $\tilde{\boldsymbol{u}}$ that corresponds to a vector $\boldsymbol{u}$ can be obtained by "lowering an index" with the metric coefficients $\eta_{\alpha\beta}$:

$$u_{\alpha} = \eta_{\alpha\beta} u^{\beta}; \text{ i.e., } \qquad u_0 = -u^0,\ u_k = u^k. \tag{2.26a}$$

Exercise 2.3. RAISING INDEX TO RECOVER THE VECTOR

One can return to the components of $\boldsymbol{u}$ by raising indices,

$$u^{\alpha} = \eta^{\alpha\beta} u_{\beta}; \tag{2.26b}$$

the matrix $\|\eta^{\alpha\beta}\|$ is defined as the inverse of $\|\eta_{\alpha\beta}\|$, and happens to equal $\|\eta_{\alpha\beta}\|$:

$$\eta^{\alpha\beta}\eta_{\beta\gamma} \equiv \delta^{\alpha}{}_{\gamma}; \qquad \eta^{\alpha\beta} = \eta_{\alpha\beta} \text{ for all } \alpha, \beta. \tag{2.27}$$

Exercise 2.4. VARIED ROUTES TO THE SCALAR PRODUCT

The scalar product of $\boldsymbol{u}$ with $\boldsymbol{v}$ can be calculated in any of the following ways:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = u^{\alpha} v^{\beta} \eta_{\alpha\beta} = u^{\alpha} v_{\alpha} = u_{\alpha} v_{\beta} \eta^{\alpha\beta}. \tag{2.28}$$

Box 2.3 DIFFERENTIALS

The "exterior derivative" or "gradient" $\boldsymbol{d}f$ of a function f is a more rigorous version of the elementary concept of "differential."

In elementary textbooks, one is presented with the differential df as representing "an infinitesimal change in the function $f(\mathscr{P})$" associated with some infinitesimal displacement of the point $\mathscr{P}$; but one will recall that the displacement of $\mathscr{P}$ is left arbitrary, albeit infinitesimal. Thus df represents a change in f in some unspecified direction.

But this is precisely what the exterior derivative $\boldsymbol{d}f$ represents. Choose a particular, infinitesimally long displacement $\boldsymbol{v}$ of the point $\mathscr{P}$. Let the displacement vector $\boldsymbol{v}$ pierce $\boldsymbol{d}f$ to give the number $\langle \boldsymbol{d}f, \boldsymbol{v}\rangle = \partial_{\boldsymbol{v}} f$. That number is the change of f in going from the tail of $\boldsymbol{v}$ to its tip. Thus $\boldsymbol{d}f$, before it has been pierced to give a number, represents the change of f in an unspecified direction. The act of piercing $\boldsymbol{d}f$ with $\boldsymbol{v}$ is the act of making explicit the direction in which the change is to be measured. The only failing of the textbook presentation, then, was its suggestion that $\boldsymbol{d}f$ was a scalar or a number; the explicit recognition of the need for specifying a direction $\boldsymbol{v}$ to reduce $\boldsymbol{d}f$ to a number $\langle \boldsymbol{d}f, \boldsymbol{v}\rangle$ shows that in fact $\boldsymbol{d}f$ is a 1-form, the gradient of f.

§2.8. THE CENTRIFUGE AND THE PHOTON

Vectors, metric, 1-forms, functions, gradients, directional derivatives: all these geometric objects and more are used in flat spacetime to represent physical quantities; and all the laws of physics must be expressible in terms of such geometric objects.

Geometric objects in action: example of centrifuge and photon

As an example, consider a high-precision redshift experiment that uses the Mössbauer effect (Figure 2.9). The emitter and the absorber of photons are attached to

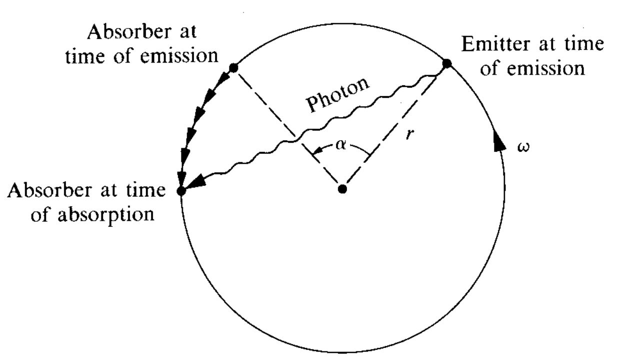


Figure 2.9.
The centrifuge and the photon.

the rim of a centrifuge at points separated by an angle α, as measured in the inertial laboratory. The emitter and absorber are at radius r as measured in the laboratory, and the centrifuge rotates with angular velocity ω. PROBLEM: What is the redshift measured,

$$z = (\lambda_{\text{absorbed}} - \lambda_{\text{emitted}})/\lambda_{\text{emitted}},$$

in terms of ω, r, and α?

SOLUTION: Let $\boldsymbol{u}_e$ be the 4-velocity of the emitter at the event of emission of a given photon; let $\boldsymbol{u}_a$ be the 4-velocity of the absorber at the event of absorption; and let $\boldsymbol{p}$ be the 4-momentum of the photon. All three quantities are vectors defined without reference to coordinates. Equally coordinate-free are the photon energies E_e and E_a measured by emitter and absorber. No coordinates are needed to describe the fact that a specific emitter emitting a specific photon attributes to it the energy E_e; and no coordinates are required in the geometric formula

$$E_e = -\boldsymbol{p} \cdot \boldsymbol{u}_e \tag{2.29}$$

for E_e. [That this formula works can be readily verified by recalling that, in the emitter's frame, $u_e{}^0 = 1$ and $u_e{}^j = 0$; so

$$E_e = -p_\alpha u_e{}^\alpha = -p_0 = +p^0$$

in accordance with the identification "(time component of 4-momentum) = (energy."] Analogous to equation (2.29) is the purely geometric formula

$$E_a = -\boldsymbol{p} \cdot \boldsymbol{u}_a$$

for the absorbed energy.

The ratio of absorbed wavelength to emitted wavelength is the inverse of the energy ratio (since $E = h\nu = hc/\lambda$):

$$\frac{\lambda_a}{\lambda_e} = \frac{E_e}{E_a} = \frac{-\boldsymbol{p} \cdot \boldsymbol{u}_e}{-\boldsymbol{p} \cdot \boldsymbol{u}_a}.$$

This ratio is most readily calculated in the inertial laboratory frame

$$\frac{\lambda_a}{\lambda_e} = \frac{p^0 u_e{}^0 - p^j u_e{}^j}{p^0 u_a{}^0 - p^j u_a{}^j} \equiv \frac{p^0 u_e{}^0 - \mathbf{p} \cdot \mathbf{u}_e}{p^0 u_a{}^0 - \mathbf{p} \cdot \mathbf{u}_a}. \tag{2.30}$$

(Here and throughout we use boldface Latin letters for three-dimensional vectors in a given Lorentz frame; and we use the usual notation and formalism of three-dimensional, Euclidean vector analysis to manipulate them.) Because the magnitude of the ordinary velocity of the rim of the centrifuge, $v = \omega r$, is unchanging in time, $u_e{}^0$ and $u_a{}^0$ are equal, and the magnitudes—but not the directions—of $\mathbf{u}_e$ and $\mathbf{u}_a$ are equal:

$$u_e{}^0 = u_a{}^0 = (1 - v^2)^{-1/2}, \; |\mathbf{u}_e| = |\mathbf{u}_a| = v/(1 - v^2)^{1/2}.$$

From the geometry of Figure 2.9, one sees that $\mathbf{u}_e$ makes the same angle with $\mathbf{p}$ as does $\mathbf{u}_a$. Consequently, $\mathbf{p} \cdot \mathbf{u}_e = \mathbf{p} \cdot \mathbf{u}_a$, and $\lambda_{\text{absorbed}}/\lambda_{\text{emitted}} = 1$. *There is no redshift!*

Notice that this solution made no reference whatsoever to Lorentz transformations—they have not even been discussed yet in this book! The power of the geometric, coordinate-free viewpoint is evident!

EXERCISES

One must have a variety of coordinate-free contacts between theory and experiment in order to use the geometric viewpoint. One such contact is the equation $E = -\boldsymbol{p} \cdot \boldsymbol{u}$ for the energy of a photon with 4-momentum $\boldsymbol{p}$, as measured by an observer with 4-velocity $\boldsymbol{u}$. Verify the following other points of contact.

Exercise 2.5. ENERGY AND VELOCITY FROM 4-MOMENTUM

A particle of rest mass m and 4-momentum $\boldsymbol{p}$ is examined by an observer with 4-velocity $\boldsymbol{u}$. Show that just as (a) the energy he measures is

$$E = -\boldsymbol{p} \cdot \boldsymbol{u}; \tag{2.31}$$

so (b) the rest mass he attributes to the particle is

$$m^2 = -\boldsymbol{p}^2; \tag{2.32}$$

(c) the momentum he measures has magnitude

$$|\boldsymbol{p}| = [(\boldsymbol{p} \cdot \boldsymbol{u})^2 + (\boldsymbol{p} \cdot \boldsymbol{p})]^{1/2}; \tag{2.33}$$

(d) the ordinary velocity $\boldsymbol{v}$ he measures has magnitude

$$|\boldsymbol{v}| = \frac{|\boldsymbol{p}|}{E}, \tag{2.34}$$

where $|\boldsymbol{p}|$ and E are as given above; and (e) the 4-vector $\boldsymbol{v}$, whose components in the observer's Lorentz frame are

$$v^0 = 0, \qquad v^j = (dx^j/dt)_{\text{for particle}} = \text{ordinary velocity},$$

is given by

$$\boldsymbol{v} = \frac{\boldsymbol{p} + (\boldsymbol{p} \cdot \boldsymbol{u})\boldsymbol{u}}{-\boldsymbol{p} \cdot \boldsymbol{u}}. \tag{2.35}$$

Exercise 2.6. TEMPERATURE GRADIENT

To each event $\mathscr{Q}$ inside the sun one attributes a temperature $T(\mathscr{Q})$, the temperature measured by a thermometer at rest in the hot gas there. Then $T(\mathscr{Q})$ is a function; no coordinates are required for its definition and discussion. A cosmic ray from outer space flies through the sun with 4-velocity $\boldsymbol{u}$. Show that, as measured by the cosmic ray's clock, the time derivative of temperature in its vicinity is

$$dT/d\tau = \partial_{\boldsymbol{u}} T = \langle \boldsymbol{d}T, \boldsymbol{u} \rangle. \tag{2.36}$$

In a local Lorentz frame inside the sun, this equation can be written

$$\frac{dT}{d\tau} = u^\alpha \frac{\partial T}{\partial x^\alpha} = \frac{1}{\sqrt{1 - \boldsymbol{v}^2}} \frac{\partial T}{\partial t} + \frac{v^j}{\sqrt{1 - \boldsymbol{v}^2}} \frac{\partial T}{\partial x^j}. \tag{2.37}$$

Why is this result reasonable?

§2.9. LORENTZ TRANSFORMATIONS

To simplify computations, one often works with the components of vectors and 1-forms, rather than with coordinate-free language. Such component manipulations sometimes involve transformations from one Lorentz frame to another. The reader is already familiar with such Lorentz transformations; but the short review in Box 2.4 will refresh his memory and acquaint him with the notation used in this book.

Lorentz transformations: of coordinates

The key entities in the Lorentz transformation are the matrices $\|\Lambda^{\alpha'}{}_{\beta}\|$ and $\|\Lambda^{\beta}{}_{\alpha'}\|$; the first transforms coordinates from an unprimed frame to a primed frame, while the second goes from primed to unprimed

$$x^{\alpha'} = \Lambda^{\alpha'}{}_{\beta} x^{\beta}, \qquad x^{\beta} = \Lambda^{\beta}{}_{\alpha'} x^{\alpha'}. \tag{2.38}$$

Since they go in opposite directions, each of the two matrices must be the inverse of the other:

$$\Lambda^{\alpha'}{}_{\beta} \Lambda^{\beta}{}_{\gamma'} = \delta^{\alpha'}{}_{\gamma'}; \qquad \Lambda^{\beta}{}_{\alpha'} \Lambda^{\alpha'}{}_{\gamma} = \delta^{\beta}{}_{\gamma}. \tag{2.39}$$

From the coordinate-independent nature of 4-velocity, $\boldsymbol{u} = (dx^{\alpha}/d\tau)\boldsymbol{e}_{\alpha}$, one readily derives the expressions

Of basis vectors

$$\boldsymbol{e}_{\alpha'} = \boldsymbol{e}_{\beta} \Lambda^{\beta}{}_{\alpha'}, \qquad \boldsymbol{e}_{\beta} = \boldsymbol{e}_{\alpha'} \Lambda^{\alpha'}{}_{\beta} \tag{2.40}$$

for the basis vectors of one frame in terms of those of the other; and from other geometric equations, such as

$$\begin{aligned} \boldsymbol{v} &= \boldsymbol{e}_{\alpha} v^{\alpha} = \boldsymbol{e}_{\beta'} v^{\beta'}, \\ \langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle &= \sigma_{\alpha} v^{\alpha} = \sigma_{\beta'} v^{\beta'}, \\ \boldsymbol{\sigma} &= \sigma_{\alpha} \boldsymbol{\omega}^{\alpha} = \sigma_{\beta'} \boldsymbol{\omega}^{\beta'}, \end{aligned}$$

one derives transformation laws

Of basis 1-forms

$$\boldsymbol{\omega}^{\alpha'} = \Lambda^{\alpha'}{}_{\beta} \boldsymbol{\omega}^{\beta}, \qquad \boldsymbol{\omega}^{\beta} = \Lambda^{\beta}{}_{\alpha'} \boldsymbol{\omega}^{\alpha'}; \tag{2.41}$$

Of components

$$v^{\alpha'} = \Lambda^{\alpha'}{}_{\beta} v^{\beta}, \qquad v^{\beta} = \Lambda^{\beta}{}_{\alpha'} v^{\alpha'}; \tag{2.42}$$

$$\sigma_{\alpha'} = \sigma_{\beta} \Lambda^{\beta}{}_{\alpha'}, \qquad \sigma_{\beta} = \sigma_{\alpha'} \Lambda^{\alpha'}{}_{\beta}. \tag{2.43}$$

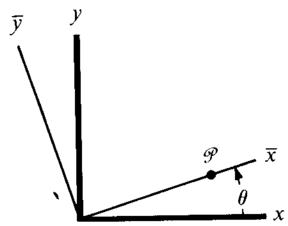
One need never memorize the index positions in these transformation laws. One need only line the indices up so that (1) free indices on each side of the equation are in the same position; and (2) summed indices appear once up and once down. Then all will be correct! (Note: the indices on Λ always run "northwest to southeast.")

Box 2.4 LORENTZ TRANSFORMATIONS

Rotation of Frame of Reference by Angle θ in x-y Plane

$$\text{Slope } s = \tan\theta; \qquad \sin\theta = \frac{s}{(1+s^2)^{1/2}}; \qquad \cos\theta = \frac{1}{(1+s^2)^{1/2}}$$

$$\begin{aligned} t &= \bar{t} \\ x &= \bar{x}\cos\theta - \bar{y}\sin\theta \\ y &= \underbrace{\bar{x}\sin\theta} + \bar{y}\cos\theta \\ z &= \bar{z} \end{aligned}$$



$$\begin{aligned} \bar{t} &= t \\ \bar{x} &= x\cos\theta + y\sin\theta \\ \bar{y} &= -x\sin\theta + y\cos\theta \\ \bar{z} &= z \end{aligned}$$

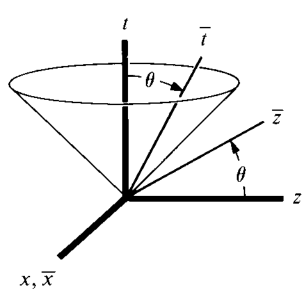
All signs follow from sign of this term. Positive by inspection of point $\mathscr{P}$.

Combination of Two Such Rotations

$$s = \frac{s_1 + s_2}{1 - s_1 s_2} \qquad \text{or} \qquad \theta = \theta_1 + \theta_2$$

Boost of Frame of Reference by Velocity Parameter α in z-t Plane

$$\text{Velocity } \beta = \tanh\alpha; \qquad \sinh\alpha = \frac{\beta}{(1-\beta^2)^{1/2}}; \qquad \cosh\alpha = \frac{1}{(1-\beta^2)^{1/2}} = \text{“}\gamma\text{”}$$



$$\begin{aligned} \tan\theta &= \text{velocity } \beta \\ &= \tanh\alpha \end{aligned}$$

$$\begin{aligned} t &= \bar{t}\cosh\alpha + \bar{z}\sinh\alpha \\ x &= \bar{x} \\ y &= \bar{y} \\ z &= \underbrace{\bar{t}\sinh\alpha} + \bar{z}\cosh\alpha \end{aligned} \qquad\qquad \begin{aligned} \bar{t} &= t\cosh\alpha - z\sinh\alpha \\ \bar{x} &= x \\ \bar{y} &= y \\ \bar{z} &= -t\sinh\alpha + z\cosh\alpha \end{aligned}$$

All signs follow from sign of this term. Positive because object at rest at $\bar{z} = 0$ in rocket frame moves in direction of increasing z in lab frame.

Matrix notation: $x^\mu = \Lambda^\mu{}_{\bar{\nu}} x^{\bar{\nu}}$, $\quad x^{\bar{\nu}} = \Lambda^{\bar{\nu}}{}_\mu x^\mu$

$$\|\Lambda^\mu{}_{\bar{\nu}}\| = \begin{Vmatrix} \cosh\alpha & 0 & 0 & \sinh\alpha \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh\alpha & 0 & 0 & \cosh\alpha \end{Vmatrix}, \quad \|\Lambda^{\bar{\nu}}{}_\mu\| = \begin{Vmatrix} \cosh\alpha & 0 & 0 & -\sinh\alpha \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\alpha & 0 & 0 & \cosh\alpha \end{Vmatrix}$$

Box 2.4 (*continued*)

Energy-momentum 4-vector	Charge density-current 4-vector
$E = \bar{E}\cosh\alpha + p^{\bar{z}}\sinh\alpha$	$\rho = \bar{\rho}\cosh\alpha + j^{\bar{z}}\sinh\alpha$
$p^x = p^{\bar{x}}$	$j^x = j^{\bar{x}}$
$p^y = p^{\bar{y}}$	$j^y = j^{\bar{y}}$
$p^z = \bar{E}\sinh\alpha + p^{\bar{z}}\cosh\alpha$	$j^z = \bar{\rho}\sinh\alpha + j^{\bar{z}}\cosh\alpha$

Aberration, incoming photon:

$$\sin\theta = \frac{-p_\perp}{E} = \frac{(1-\beta^2)^{1/2}\sin\bar{\theta}}{1-\beta\cos\bar{\theta}} \qquad \sin\bar{\theta} = \frac{-\bar{p}_\perp}{\bar{E}} = \frac{(1-\beta^2)^{1/2}\sin\theta}{1+\beta\cos\theta}$$

$$\cos\theta = \frac{-p^z}{E} = \frac{\cos\bar{\theta}-\beta}{1-\beta\cos\bar{\theta}} \qquad \cos\bar{\theta} = \frac{-\bar{p}^{\bar{z}}}{\bar{E}} = \frac{\cos\theta+\beta}{1+\beta\cos\theta}$$

$$\tan(\theta/2) = e^{\alpha}\tan(\bar{\theta}/2) \qquad \tan(\bar{\theta}/2) = e^{-\alpha}\tan(\theta/2)$$

Combination of Two Boosts in Same Direction

$$\beta = \frac{\beta_1+\beta_2}{1+\beta_1\beta_2} \quad \text{or} \quad \alpha = \alpha_1 + \alpha_2.$$

General Combinations of Boosts and Rotations

Spinor formalism of Chapter 41

Poincaré Transformation

$$x^\mu = \Lambda^\mu{}_{\alpha'}x^{\alpha'} + a^\mu.$$

Condition on the Lorentz part of this transformation:

$$ds'^2 = \eta_{\alpha'\beta'}\,dx^{\alpha'}\,dx^{\beta'} = ds^2 = \eta_{\mu\nu}\Lambda^\mu{}_{\alpha'}\Lambda^\nu{}_{\beta'}\,dx^{\alpha'}\,dx^{\beta'}$$

or $\Lambda^T\eta\Lambda = \eta$ (matrix equation, with T indicating "transposed," or rows and columns interchanged).

Effect of transformation on other quantities:

$u^\mu = \Lambda^\mu{}_{\alpha'}u^{\alpha'}$	(4-velocity)	$u_{\alpha'} = u_\mu\Lambda^\mu{}_{\alpha'}$;
$p^\mu = \Lambda^\mu{}_{\alpha'}p^{\alpha'}$	(4-momentum)	$p_{\alpha'} = p_\mu\Lambda^\mu{}_{\alpha'}$;
$F^{\mu\nu} = \Lambda^\mu{}_{\alpha'}\Lambda^\nu{}_{\beta'}F^{\alpha'\beta'}$	(electromagnetic field)	$F_{\alpha'\beta'} = F_{\mu\nu}\Lambda^\mu{}_{\alpha'}\Lambda^\nu{}_{\beta'}$;
$\boldsymbol{e}_{\alpha'} = \boldsymbol{e}_\mu\Lambda^\mu{}_{\alpha'}$	(basis vectors);	
$\boldsymbol{\omega}^{\alpha'} = \Lambda^{\alpha'}{}_\mu\boldsymbol{\omega}^\mu$	(basis 1-forms);	
$\boldsymbol{u} = \boldsymbol{e}_{\alpha'}u^{\alpha'} = \boldsymbol{e}_\mu u^\mu = \boldsymbol{u}$	(the 4-velocity *vector*).	

Exercise 2.7. BOOST IN AN ARBITRARY DIRECTION **EXERCISE**

An especially useful Lorentz transformation has the matrix components

$$\begin{aligned}
\Lambda^{0'}{}_{0} &= \gamma \equiv \frac{1}{\sqrt{1-\beta^2}},\\
\Lambda^{0'}{}_{j} &= \Lambda^{j'}{}_{0} = -\beta\gamma n^j,\\
\Lambda^{j'}{}_{k} &= \Lambda^{k'}{}_{j} = (\gamma - 1)n^j n^k + \delta^{jk},\\
\Lambda^{\mu}{}_{\nu'} &= (\text{same as } \Lambda^{\nu'}{}_{\mu} \text{ but with } \beta \text{ replaced by } -\beta),
\end{aligned} \tag{2.44}$$

where β, n^1, n^2, and n^3 are parameters, and $\boldsymbol{n}^2 \equiv (n^1)^2 + (n^2)^2 + (n^3)^2 = 1$. Show (a) that this does satisfy the condition $\Lambda^T \eta \Lambda = \eta$ required of a Lorentz transformation (see Box 2.4); (b) that the primed frame moves with ordinary velocity $\beta\boldsymbol{n}$ as seen in the unprimed frame; (c) that the unprimed frame moves with ordinary velocity $-\beta\boldsymbol{n}$ (i.e., $v^{1'} = -\beta n^1$, $v^{2'} = -\beta n^2$, $v^{3'} = -\beta n^3$) as seen in the primed frame; and (d) that for motion in the z direction, the transformation matrices reduce to the familiar form

$$\|\Lambda^{\nu'}{}_{\mu}\| = \begin{Vmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{Vmatrix}, \qquad \|\Lambda^{\mu}{}_{\nu'}\| = \begin{Vmatrix} \gamma & 0 & 0 & \beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta\gamma & 0 & 0 & \gamma \end{Vmatrix}. \tag{2.45}$$

§2.10. COLLISIONS

Whatever the physical entity, whether it is an individual mass in motion, or a torrent of fluid, or a field of force, or the geometry of space itself, it is described in classical general relativity as a geometric object of its own characteristic kind. Each such object is built directly or by abstraction from identifiable points, and needs no coordinates for its representation. It has been seen how this coordinate-free description translates into, and how it can be translated out of, the language of coordinates and components, and how components in a local Lorentz frame transform under a Lorentz transformation. Turn now to two elementary applications of this mathematical machinery to a mass in motion. One has to do with short-range forces (collisions, this section); the other, with the long-range electromagnetic force (Lorentz force law, next chapter).

Scattering of particles

In a collision, all the change in momentum is concentrated in a time that is short compared to the time of observation. Moreover, the target is typically so small, and quantum mechanics so dominating, that a probabilistic description is the right one. A quantity

$$d\sigma = \left(\frac{d\sigma}{d\Omega}\right)_\theta d\Omega \tag{2.46}$$

gives the cross section (cm^2) for scattering into the element of solid angle $d\Omega$ at the deflection angle θ; a more complicated expression gives the probability that the original particle will enter the aperture $d\Omega$ at a given polar angle θ and azimuth ϕ and with energy E to $E + dE$, while simultaneously products of reaction also emerge into specified energy intervals and into specified angular apertures. It would be out of place here to enter into the calculation of such cross sections, though it is a fascinating branch of atomic physics. It is enough to note that the cross section is an area oriented perpendicular to the line of travel of the incident particle. Therefore it is unaffected by any boost of the observer in that direction, provided of course that energies and angles of emergence of the particles are transformed in accordance with the magnitude of that boost ("same events seen in an altered reference system").

Conservation of energy-momentum in a collision

Over and above any such detailed account of the encounter as follows from the local dynamic analysis, there stands the law of conservation of energy-momentum:

$$\sum_{\substack{\text{original} \\ \text{particles, } J}} \boldsymbol{p}_J = \sum_{\substack{\text{final} \\ \text{particles, } K}} \boldsymbol{p}_K \,. \tag{2.47}$$

Out of this relation, one wins without further analysis such simple results as the following. (1) A photon traveling as a plane wave through empty space cannot split (not true for a focused photon!). (2) When a high-energy electron strikes an electron at rest in an elastic encounter, and the two happen to come off sharing the energy equally, then the angle between their directions of travel is less than the Newtonian value of 90°, and the deficit gives a simple measure of the energy of the primary. (3) When an electron makes a head-on elastic encounter with a proton, the formula for the fraction of kinetic energy transferred has three rather different limiting forms, according to whether the energy of the electron is nonrelativistic, relativistic, or extreme-relativistic. (4) The threshold for the production of an (e^+, e^-) pair by a photon in the field of force of a massive nucleus is $2m_e$. (5) The threshold for the production of an (e^+, e^-) pair by a photon in an encounter with an electron at rest is $4m_e$ (or $4m_e - \epsilon$ when account is taken of the binding of the $e^+e^-e^-$ system in a very light "molecule"). All these results (topics for independent projects!) and more can be read out of the law of conservation of energy-momentum. For more on this topic, see Blaton (1950), Hagedorn (1964), and Chapter 4 and the last part of Chapter 5 of Sard (1970).

CHAPTER 3

THE ELECTROMAGNETIC FIELD

The rotating armatures of every generator and every motor in this age of electricity are steadily proclaiming the truth of the relativity theory to all who have ears to hear.

LEIGH PAGE (1941)

§3.1. THE LORENTZ FORCE AND THE ELECTROMAGNETIC FIELD TENSOR

At the opposite extreme from an impulsive change of momentum in a collision (the last topic of Chapter 2) is the gradual change in the momentum of a charged particle under the action of electric and magnetic forces (the topic treated here).

Lorentz force as definer of fields and predicter of motions

Let electric and magnetic fields act on a system of charged particles. The accelerations of the particles reveal the electric and magnetic field strengths. In other words, the Lorentz force law, plus measurements on the components of acceleration of test particles, can be viewed as defining the components of the electric and magnetic fields. Once the field components are known from the accelerations of a few test particles, they can be used to predict the accelerations of other test particles (Box 3.1). Thus the Lorentz force law does double service (1) as definer of fields and (2) as predicter of motions.

*Here and elsewhere in science, as stressed not least by Henri Poincaré, that view is out of date which used to say, "Define your terms before you proceed." All the laws and theories of physics, including the Lorentz force law, have this deep and subtle character, that they both define the concepts they use (here **B** and **E**) and make statements about these concepts. Contrariwise, the absence of some body of theory, law, and principle deprives one of the means properly to define or even to use concepts. Any forward step in human knowledge is truly creative in this sense: that theory, concept, law, and method of measurement—forever inseparable—are born into the world in union.*

Box 3.1 LORENTZ FORCE LAW AS BOTH DEFINER OF FIELDS AND PREDICTER OF MOTIONS

How one goes about determining the components of the field from measurements of accelerations is not different in principle for electromagnetism and for gravitation. Compare the equations in the two cases:

$$\frac{d^2x^\alpha}{d\tau^2} = \frac{e}{m} F^\alpha{}_\beta u^\beta \text{ in a Lorentz frame,} \tag{1}$$

and

$$\frac{D^2\xi^\alpha}{d\tau^2} = -R^\alpha{}_{\beta\gamma\delta} u^\beta \xi^\gamma u^\delta \text{ in any coordinate system.} \tag{2}$$

To make explicit the simpler procedure for electromagnetism will indicate in broad outline how one similarly determines all the components of $R^\alpha{}_{\beta\gamma\delta}$ for gravity. Begin by asking how many test particles one needs to determine the three components of $\boldsymbol{B}$ and the three components of $\boldsymbol{E}$ in the neighborhood under study. For one particle, three components of acceleration are measurable; for a second particle, three more. Enough? No! The information from the one duplicates in part the information from the other. The proof? Whatever the state of motion of the first test particle, pick one's Lorentz frame to be moving the same way. Having zero velocity in this frame, the particle has a zero response to any magnetic field. The electric field alone acts on the particle. The three components of its acceleration give directly the three components E_x, E_y, E_z of the electric field. The second test particle cannot be at rest if it is to do more than duplicate the information provided by the first test particle. Orient the x-axis of the frame of reference parallel to the direction of motion of this second particle, which will then respond to and measure the components B_y and B_z of the magnetic field. Not so B_x! The acceleration in the x-direction merely remeasures the already once measured E_x. To evaluate B_x, a third test particle is required, but it then gives duplicate information about the other field components. The alternative? Use all N particles simultaneously and on the same democratic footing, both in the evaluation of the six $F_{\alpha\beta}$ and in the testing of the Lorentz force, by applying the method of least squares. Thus, write the discrepancy between predicted and observed acceleration of the Kth particle in the form

$$\dot{u}_\alpha{}^K - \frac{e}{m} F_{\alpha\beta} u^{\beta,K} = \delta a_\alpha{}^K. \tag{3}$$

Take the squared magnitude of this discrepancy and sum over all the particles

$$S = \sum_k \eta^{\alpha\beta} \delta a_\alpha{}^K \delta a^K{}_\beta. \tag{4}$$

In this expression, everything is regarded as known except the six $F_{\alpha\beta}$. Minimize with respect to these six unknowns. In this way, arrive at six equations for the components of $\boldsymbol{B}$ and $\boldsymbol{E}$. These equations once solved, one goes back to (3) to test the Lorentz force law.

The 6×6 determinant of the coefficients in the equation for the $F_{\alpha\beta}$ automatically vanishes when there are only two test particles. The same line of reasoning permits one to determine the minimum number of test particles required to determine all the components of the Riemann curvature tensor.

The Lorentz force law, written in familiar three-dimensional notation, with $\boldsymbol{E}$ = electric field, $\boldsymbol{B}$ = magnetic field, $\boldsymbol{v}$ = ordinary velocity of particle, $\boldsymbol{p}$ = momentum of particle, e = charge of particle, reads

The three-dimensional version of the Lorentz force law

$$(d\boldsymbol{p}/dt) = e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}). \tag{3.1}$$

Useful though this version of the equation may be, it is far from the geometric spirit of Einstein. A fully geometric equation will involve the test particle's energy-momentum 4-vector, $\boldsymbol{p}$, not just the spatial part $\boldsymbol{p}$ as measured in a specific Lorentz frame; and it will ask for the rate of change of momentum not as measured by a specific Lorentz observer (d/dt), but as measured by the only clock present *à priori* in the problem: the test particle's own clock ($d/d\tau$). Thus, the lefthand side of a fully geometric equation will read

$$d\boldsymbol{p}/d\tau = .$$

The righthand side, the Lorentz 4-force, must also be a frame-independent object. It will be linear in the particle's 4-velocity $\boldsymbol{u}$, since the frame-dependent expression

$$\frac{d\boldsymbol{p}}{d\tau} = \frac{1}{\sqrt{1-\boldsymbol{v}^2}}\frac{d\boldsymbol{p}}{dt} = \frac{e}{\sqrt{1-\boldsymbol{v}^2}}(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) = e(u^0\boldsymbol{E} + \boldsymbol{u} \times \boldsymbol{B}) \tag{3.2a}$$

is linear in the components of $\boldsymbol{u}$. Consequently, there must be a linear machine named ***Faraday***, or $\boldsymbol{F}$, or "electromagnetic field tensor," with a slot into which one inserts the 4-velocity of a test particle. The output of this machine, multiplied by the particle's charge, must be the electromagnetic 4-force that it feels:

Electromagnetic field tensor defined

Geometrical version of Lorentz force law

$$d\boldsymbol{p}/d\tau = e\boldsymbol{F}(\boldsymbol{u}). \tag{3.3}$$

By comparing this geometric equation with the original Lorentz force law (equation 3.2a), and with the companion energy-change law

$$\frac{dp^0}{d\tau} = \frac{1}{\sqrt{1-\boldsymbol{v}^2}}\frac{dE}{dt} = \frac{1}{\sqrt{1-\boldsymbol{v}^2}}e\boldsymbol{E}\cdot\boldsymbol{v} = e\boldsymbol{E}\cdot\boldsymbol{u}, \tag{3.2b}$$

one can read off the components of $\boldsymbol{F}$ in a specific Lorentz frame. The components of $d\boldsymbol{p}/d\tau$ are $dp^\alpha/d\tau$, and the components of $e\boldsymbol{F}(\boldsymbol{u})$ can be written (definition of $F^\alpha{}_\beta$!) $eF^\alpha{}_\beta u^\beta$. Consequently

$$dp^\alpha/d\tau = eF^\alpha{}_\beta u^\beta \tag{3.4}$$

must reduce to equations (3.2a,b). Indeed it does if one makes the identification

Components of electromagnetic field tensor

$$\|F^\alpha{}_\beta\| = \begin{array}{c} \\ \alpha=0 \\ \alpha=1 \\ \alpha=2 \\ \alpha=3 \end{array} \begin{array}{c} \begin{array}{cccc} \beta=0 & \beta=1 & \beta=2 & \beta=3 \end{array} \\ \left\| \begin{array}{cccc} 0 & E_x & E_y & E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{array} \right\| \end{array}. \tag{3.5}$$

More often seen in the literature are the "covariant components," obtained by lowering an index with the metric components:

$$F_{\alpha\beta} = \eta_{\alpha\gamma} F^{\gamma}{}_{\beta}; \tag{3.6}$$

$$\|F_{\alpha\beta}\| = \left\| \begin{matrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{matrix} \right\|. \tag{3.7}$$

This matrix equation demonstrates the unity of the electric and magnetic fields. Neither one by itself, $\boldsymbol{E}$ or $\boldsymbol{B}$, is a frame-independent, geometric entity. But merged together into a single entity, $\boldsymbol{F}$ = ***Faraday***, they acquire a meaning and significance that transcends coordinates and reference frames.

EXERCISE

Exercise 3.1.

Derive equations (3.5) and (3.7) for the components of ***Faraday*** by comparing (3.4) with (3.2a,b), and by using definition (3.6).

---

§3.2. TENSORS IN ALL GENERALITY

Examples of tensors

A digression is in order. Now on the scene are several different tensors: the metric tensor $\boldsymbol{g}$ (§2.4), the Riemann curvature tensor ***Riemann*** (§1.6), the electromagnetic field tensor ***Faraday*** (§3.1). Each has been defined as a linear machine with input slots for vectors, and with an output that is either a real number, e.g., $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v})$, or a vector, e.g., ***Riemann*** $(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ and ***Faraday*** $(\boldsymbol{u})$.

Should one make a distinction between tensors whose outputs are scalars, and tensors whose outputs are vectors? No! A tensor whose output is a vector can be reinterpreted trivially as one whose output is a scalar. Take, for example, ***Faraday*** = $\boldsymbol{F}$. Add a new slot for the insertion of an arbitrary 1-form $\boldsymbol{\sigma}$, and gears and wheels that guarantee the output

$$\boldsymbol{F}(\boldsymbol{\sigma}, \boldsymbol{u}) = \langle \boldsymbol{\sigma}, \boldsymbol{F}(\boldsymbol{u}) \rangle = \text{real number}. \tag{3.8}$$

Then permit the user to choose whether he inserts only a vector, and gets out the vector $\boldsymbol{F}(\ldots, \boldsymbol{u}) = \boldsymbol{F}(\boldsymbol{u})$, or whether he inserts a form and a vector, and gets out the number $\boldsymbol{F}(\boldsymbol{\sigma}, \boldsymbol{u})$. The same machine will do both jobs. Moreover, in terms of components in a given Lorentz frame, both jobs are achieved very simply:

$$\begin{aligned} &\boldsymbol{F}(\ldots, \boldsymbol{u}) \text{ is a vector with components } F^{\alpha}{}_{\beta} u^{\beta}; \\ &\boldsymbol{F}(\boldsymbol{\sigma}, \boldsymbol{u}) \text{ is the number } \langle \boldsymbol{\sigma}, \boldsymbol{F}(\ldots, \boldsymbol{u}) \rangle = \sigma_{\alpha} F^{\alpha}{}_{\beta} u^{\beta}. \end{aligned} \tag{3.9}$$

By analogy, one defines the most general tensor $\boldsymbol{H}$ and its rank $\binom{n}{m}$ as follows: $\boldsymbol{H}$ is a linear machine with n input slots for n 1-forms, and m input slots for m vectors; given the requested input, it puts out a real number denoted

Definition of tensor as linear machine that converts vectors and 1-forms into numbers

$$\boldsymbol{H}(\underbrace{\boldsymbol{\sigma}, \boldsymbol{\lambda}, \ldots, \boldsymbol{\beta}}_{n \text{ 1-forms}}, \underbrace{\boldsymbol{u}, \boldsymbol{v}, \ldots, \boldsymbol{w}}_{m \text{ vectors}}). \tag{3.10}$$

For most tensors the output changes when two input vectors are interchanged,

$$\boldsymbol{Riemann}(\boldsymbol{\sigma}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \neq \boldsymbol{Riemann}(\boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}), \tag{3.11}$$

or when two input 1-forms are interchanged.

Choose a specific tensor $\boldsymbol{S}$, of rank $\binom{2}{1}$ for explicitness. Into the slots of $\boldsymbol{S}$, insert the basis vectors and 1-forms of a specific Lorentz coordinate frame. The output is a "component of $\boldsymbol{S}$ in that frame":

$$S^{\alpha\beta}{}_{\gamma} \equiv \boldsymbol{S}(\boldsymbol{\omega}^{\alpha}, \boldsymbol{\omega}^{\beta}, \boldsymbol{e}_{\gamma}). \tag{3.12}$$

Components of a tensor

This defines components. Knowing the components in a specific frame, one can easily calculate the output produced from any input forms and vectors:

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{v}) &= \boldsymbol{S}(\sigma_{\alpha}\boldsymbol{\omega}^{\alpha}, \rho_{\beta}\boldsymbol{\omega}^{\beta}, v^{\gamma}\boldsymbol{e}_{\gamma}) = \sigma_{\alpha}\rho_{\beta}v^{\gamma}\boldsymbol{S}(\boldsymbol{\omega}^{\alpha}, \boldsymbol{\omega}^{\beta}, \boldsymbol{e}_{\gamma}) \\ &= S^{\alpha\beta}{}_{\gamma}\sigma_{\alpha}\rho_{\beta}v^{\gamma}. \end{aligned} \tag{3.13}$$

Tensor's machine action expressed in terms of components

And knowing the components of $\boldsymbol{S}$ in one Lorentz frame (unprimed), plus the Lorentz transformation matrices $\|\Lambda^{\alpha'}{}_{\beta}\|$ and $\|\Lambda^{\beta}{}_{\alpha'}\|$ which link that frame with another (primed), one can calculate the components in the new (primed) frame. As shown in exercise 3.2, one need only apply a matrix to each index of $\boldsymbol{S}$, lining up the matrix indices in the logical manner

$$S^{\mu'\nu'}{}_{\lambda'} = S^{\alpha\beta}{}_{\gamma}\Lambda^{\mu'}{}_{\alpha}\Lambda^{\nu'}{}_{\beta}\Lambda^{\gamma}{}_{\lambda'}. \tag{3.14}$$

Lorentz transformation of components

A slight change of the internal gears and wheels inside the tensor enables one of its 1-form slots to accept a vector. All that is necessary is a mechanism to convert an input vector $\boldsymbol{n}$ into its corresponding 1-form $\tilde{\boldsymbol{n}}$ and then to put that 1-form into the old machinery. Thus, denoting the modified tensor by the same symbol $\boldsymbol{S}$ as was used for the original tensor, one demands

$$\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{n}, \boldsymbol{v}) = \boldsymbol{S}(\boldsymbol{\sigma}, \tilde{\boldsymbol{n}}, \boldsymbol{v}); \tag{3.15}$$

Modifying a tensor to accept either a vector or a 1-form into each slot

or, in component notation

$$S^{\alpha}{}_{\beta\gamma}\sigma_{\alpha}n^{\beta}v^{\gamma} = S^{\alpha\beta}{}_{\gamma}\sigma_{\alpha}n_{\beta}v^{\gamma}. \tag{3.15$'$}$$

This is achieved if one raises and lowers the indices of $\boldsymbol{S}$ using the components of the metric:

$$S^{\alpha}{}_{\beta\gamma} = \eta_{\beta\mu}S^{\alpha\mu}{}_{\gamma}, \qquad S^{\alpha\mu}{}_{\gamma} = \eta^{\mu\beta}S^{\alpha}{}_{\beta\gamma}. \tag{3.16}$$

Raising and lowering indices

(See exercise 3.3 below.) By using the same symbol $\boldsymbol{S}$ for the original tensor and the modified tensor, one allows each slot to accept either a 1-form or a vector, so one loses sight of whether $\boldsymbol{S}$ is a $\binom{2}{1}$ tensor, or a $\binom{1}{2}$ tensor, or a $\binom{3}{0}$ tensor, or a $\binom{0}{3}$ tensor; one only distinguishes its total rank, 3. *Terminology:* an "upstairs index" is called "contravariant"; a "downstairs" index is called "covariant." Thus in $S^{\alpha}{}_{\beta\gamma}$, "$\alpha$" is a contravariant index, while "β" and "γ" are covariant indices.

Because tensors are nothing but functions, they can be added (if they have the same rank!) and multiplied by numbers in the usual way: the output of the rank-3 tensor $a\boldsymbol{S} + b\boldsymbol{Q}$, when vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ are put in, is

$$(a\boldsymbol{S} + b\boldsymbol{Q})(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \equiv a\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) + b\boldsymbol{Q}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}). \tag{3.17}$$

Several other important operations on tensors are explored in the following exercises. They and the results of the exercises will be used freely in the material that follows.

EXERCISES

Exercise 3.2. TRANSFORMATION LAW FOR COMPONENTS OF A TENSOR

From the transformation laws for components of vectors and 1-forms, derive the transformation law (3.14).

Exercise 3.3. RAISING AND LOWERING INDICES

Derive equations (3.16) from equation (3.15′) plus the law $n_\alpha = \eta_{\alpha\beta} n^\beta$ for getting the components of the 1-form $\tilde{\boldsymbol{n}}$ from the components of its corresponding vector $\boldsymbol{n}$.

Exercise 3.4. TENSOR PRODUCT

Given any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, one defines the second-rank tensor $\boldsymbol{u} \otimes \boldsymbol{v}$ ("tensor product of $\boldsymbol{u}$ with $\boldsymbol{v}$") to be a machine, with two input slots, whose output is the number

$$(\boldsymbol{u} \otimes \boldsymbol{v})(\boldsymbol{\sigma}, \boldsymbol{\lambda}) = \langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \langle \boldsymbol{\lambda}, \boldsymbol{v} \rangle \tag{3.18}$$

when 1-forms $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ are inserted. Show that the components of $\boldsymbol{T} = \boldsymbol{u} \otimes \boldsymbol{v}$ are the products of the components of $\boldsymbol{u}$ and $\boldsymbol{v}$:

$$T^{\alpha\beta} = u^\alpha v^\beta, \qquad T_\alpha{}^\beta = u_\alpha v^\beta, \qquad T_{\alpha\beta} = u_\alpha v_\beta. \tag{3.19}$$

Extend the definition to several vectors and forms,

$$(\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{\beta} \otimes \boldsymbol{w})(\boldsymbol{\sigma}, \boldsymbol{\lambda}, \boldsymbol{n}, \boldsymbol{\zeta}) = \langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \langle \boldsymbol{\lambda}, \boldsymbol{v} \rangle \langle \boldsymbol{\beta}, \boldsymbol{n} \rangle \langle \boldsymbol{\zeta}, \boldsymbol{w} \rangle, \tag{3.20}$$

and show that the product rule for components still holds:

$$\begin{gathered} \boldsymbol{S} = \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{\beta} \otimes \boldsymbol{w} \text{ has components} \\ S^{\mu\nu}{}_\lambda{}^\xi = u^\mu v^\nu \beta_\lambda w^\xi. \end{gathered} \tag{3.21}$$

Exercise 3.5. BASIS TENSORS

Show that a tensor $\boldsymbol{M}$ with components $M^{\alpha\beta}{}_\gamma{}^\delta$ in a given Lorentz frame can be reconstructed from its components and from the basis 1-forms and vectors of that frame as follows:

$$\boldsymbol{M} = M^{\alpha\beta}{}_\gamma{}^\delta \boldsymbol{e}_\alpha \otimes \boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\gamma \otimes \boldsymbol{e}_\delta. \tag{3.22}$$

(For a special case of this, see Box 3.2.)

Box 3.2 THE METRIC IN DIFFERENT LANGUAGES

A. Geometric Language

$\boldsymbol{g}$ is a linear, symmetric machine with two slots for insertion of vectors. When vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are inserted, the output of $\boldsymbol{g}$ is their scalar product:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}.$$

B. Component Language

$\eta_{\mu\nu}$ are the metric components. They are used to calculate the scalar product of two vectors from components in a specific Lorentz frame:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \eta_{\mu\nu} u^\mu v^\nu.$$

C. Coordinate-Based Geometric Language

The metric $\boldsymbol{g}$ can be written, in terms of basis 1-forms of a specific Lorentz frame, as

$$\boldsymbol{g} = \eta_{\mu\nu} \boldsymbol{\omega}^\mu \otimes \boldsymbol{\omega}^\nu = \eta_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$$

[see equations (2.18) and (3.22)].

D. Connection to the Elementary Concept of Line Element

Box 2.3 demonstrated the correspondence between the gradient $\boldsymbol{d}f$ of a function, and the elementary concept df of a differential change of f in some unspecified direction. There is a similar correspondence between the metric, written as $\eta_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$, and the elementary concept of "line element," written as $ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu$. This elementary line element, as expounded in many special relativity texts, represents the squared length of the displacement "dx^μ" in an unspecified direction. The metric $\eta_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$ does the same. Pick a specific infinitesimal displacement vector $\boldsymbol{\xi}$, and insert it into the slots of $\eta_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$. The output will be $\boldsymbol{\xi}^2 = \eta_{\mu\nu}\xi^\mu\xi^\nu$, the squared length of the displacement. Before $\boldsymbol{\xi}$ is inserted, $\eta_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$ has the potential to tell the squared length of any vector; the insertion of $\boldsymbol{\xi}$ converts potentiality into actuality: the numerical value of $\boldsymbol{\xi}^2$.

Because the metric $\eta_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$ and the line element $ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu$ perform this same function of representing the squared length of an unspecified infinitesimal displacement, there is no conceptual distinction between them. One sometimes uses the symbols $\boldsymbol{ds^2}$ to denote the metric; one sometimes gets pressed and writes it as $\boldsymbol{ds^2} = \eta_{\mu\nu}\, \boldsymbol{d}x^\mu\, \boldsymbol{d}x^\nu$, omitting the "$\otimes$"; and one sometimes even gets so pressed as to use nonbold characters, so that no notational distinction remains at all between metric and elementary line element:

$$\boldsymbol{g} = \boldsymbol{ds^2} = ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu.$$

Exercise 3.6. *Faraday* MACHINERY AT WORK

An observer with 4-velocity $\boldsymbol{u}$ picks out three directions in spacetime that are orthogonal and purely spatial (no time part) as seen in his frame. Let $\boldsymbol{e}_{\hat{1}}, \boldsymbol{e}_{\hat{2}}, \boldsymbol{e}_{\hat{3}}$ be unit vectors in those directions and let them be oriented in a righthanded way ($\boldsymbol{e}_{\hat{1}} \cdot \boldsymbol{e}_{\hat{2}} \times \boldsymbol{e}_{\hat{3}} = +1$ in three-dimensional language). Why do the following relations hold?

$$\boldsymbol{e}_{\hat{j}} \cdot \boldsymbol{u} = 0, \qquad \boldsymbol{e}_{\hat{j}} \cdot \boldsymbol{e}_{\hat{k}} = \delta_{jk}.$$

What vectors are to be inserted in the two slots of the electromagnetic field tensor ***Faraday*** if one wants to get out the electric field along $\boldsymbol{e}_{\hat{j}}$ as measured by this observer? What vectors must be inserted to get the magnetic field he measures along $\boldsymbol{e}_{\hat{j}}$?

§3.3. THREE-PLUS-ONE VIEW VERSUS GEOMETRIC VIEW

The power of the geometric view of physics

Great computational and conceptual power resides in Einstein's geometric view of physics. Ideas that seem complex when viewed in the everyday "space-plus-time" or "3 + 1" manner become elegant and simple when viewed as relations between geometric objects in four-dimensional spacetime. Derivations that are difficult in 3 + 1 language simplify in geometric language.

Example of electromagnetism

The electromagnetic field is a good example. In geometric language, it is described by a second-rank, antisymmetric tensor ("2-form") $\boldsymbol{F}$, which requires no coordinates for its definition. This tensor produces a 4-force on any charged particle given by

$$d\boldsymbol{p}/d\tau = e\boldsymbol{F}(\boldsymbol{u}).$$

It is all so simple!

By contrast, consider the "3 + 1" viewpoint. In a given Lorentz frame, there is an electric field $\boldsymbol{E}$ and a magnetic field $\boldsymbol{B}$. They push on a particle in accordance with

$$d\boldsymbol{p}/dt = e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}).$$

Transformation law for electric and magnetic fields

But the values of $\boldsymbol{p}$, $\boldsymbol{E}$, $\boldsymbol{v}$, and $\boldsymbol{B}$ all change when one passes from the given Lorentz frame to a new one. For example, the electric and magnetic fields viewed from a rocket ship ("barred" frame) are related to those viewed in the laboratory ("unbarred" frame) by

$$\begin{aligned} \bar{\boldsymbol{E}}_{\|} &= \boldsymbol{E}_{\|}, \qquad \bar{\boldsymbol{E}}_{\perp} = \frac{1}{\sqrt{1-\beta^2}}(\boldsymbol{E}_{\perp} + \boldsymbol{\beta} \times \boldsymbol{B}_{\perp}), \\ \bar{\boldsymbol{B}}_{\|} &= \boldsymbol{B}_{\|}, \qquad \bar{\boldsymbol{B}}_{\perp} = \frac{1}{\sqrt{1-\beta^2}}(\boldsymbol{B}_{\perp} - \boldsymbol{\beta} \times \boldsymbol{E}_{\perp}). \end{aligned} \tag{3.23}$$

(Here "$\|$" means component along direction of rocket's motion; "$\perp$" means perpendicular component; and $\beta^j = dx^j{}_{\text{rocket}}/dt$ is the rocket's ordinary velocity.) The analogous transformation laws for the particle's momentum $\boldsymbol{p}$ and ordinary velocity

$\boldsymbol{v}$, and for the coordinate time t, all conspire—as if by magic, it seems, from the $3 + 1$ viewpoint—to maintain the validity of the Lorentz force law in all frames.

Not only is the geometric view far simpler than the $3 + 1$ view, it can even derive the $3 + 1$ equations with greater ease than can the $3 + 1$ view itself. Consider, for example, the transformation law (3.23) for the electric and magnetic fields. The geometric view derives it as follows: (1) Orient the axes of the two frames so their relative motion is in the z-direction. (2) Perform a simple Lorentz transformation (equation 2.45) on the components of the electromagnetic field tensor:

$$\begin{aligned} \bar{E}_{\|} = \bar{E}_z = F_{\bar{3}\bar{0}} &= \Lambda^{\alpha}{}_{\bar{3}}\Lambda^{\beta}{}_{\bar{0}}F_{\alpha\beta} = \gamma^2 F_{30} + \beta^2\gamma^2 F_{03} \\ &= (1-\beta^2)\gamma^2 F_{30} = F_{30} = E_z = E_{\|}, \\ \bar{E}_x = F_{\bar{1}\bar{0}} &= \Lambda^{\alpha}{}_{\bar{1}}\Lambda^{\beta}{}_{\bar{0}}F_{\alpha\beta} = \gamma F_{10} + \beta\gamma F_{13} = \gamma(E_x - \beta B_y), \end{aligned} \tag{3.24}$$

etc.

By contrast, the $3 + 1$ view shows much more work. A standard approach is based on the Lorentz force law and energy-change law (3.2a,b), written in the slightly modified form

$$m\frac{d^2\bar{x}}{d\tau^2} = e\left(\bar{E}_x\frac{d\bar{t}}{d\tau} + 0\frac{d\bar{x}}{d\tau} + \bar{B}_z\frac{d\bar{y}}{d\tau} - \bar{B}_y\frac{d\bar{z}}{d\tau}\right), \tag{3.25}$$

. . . (three additional equations) . . . .

It proceeds as follows (details omitted because of their great length!):

(1) Substitute for the $d^2\bar{x}/d\tau^2$, etc., the expression for these quantities in terms of the $d^2x/d\tau^2, \ldots$ (Lorentz transformation).

(2) Substitute for the $d^2x/d\tau^2, \ldots$ the expression for these accelerations in terms of the laboratory $\boldsymbol{E}$ and $\boldsymbol{B}$ (Lorentz force law).

(3) In these expressions, wherever the components $dx/d\tau$ of the 4-velocity in the laboratory frame appear, substitute expressions in terms of the 4-velocities in the rocket frame (inverse Lorentz transformation).

(4) In (3.25) as thus transformed, demand equality of left and right sides for all values of the $d\bar{x}/d\tau$, etc. (validity for all test particles).

(5) In this way arrive at the expressions (3.23) for the $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$ in terms of the $\boldsymbol{E}$ and $\boldsymbol{B}$.

The contrast in difficulty is obvious!

§3.4. MAXWELL'S EQUATIONS

Turn now from the action of the field on a charge, and ask about the action of a charge on the field, or, more generally, ask about the dynamics of the electromagnetic field, charge or no charge. Begin with the simplest of Maxwell's equations in a specific Lorentz frame, the one that says there are no free magnetic poles:

Magnetodynamics derived from magnetostatics

$$\nabla \cdot \boldsymbol{B} \equiv \operatorname{div} \boldsymbol{B} = \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0. \tag{3.26}$$

This statement has to be true in all Lorentz frames. It is therefore true in the rocket frame:

$$\frac{\partial \bar{B}_x}{\partial \bar{x}} + \frac{\partial \bar{B}_y}{\partial \bar{y}} + \frac{\partial \bar{B}_z}{\partial \bar{z}} = 0. \tag{3.27}$$

For an infinitesimal Lorentz transformation in the x-direction (nonrelativistic velocity β), one has (see Box 2.4 and equations 3.23)

$$\bar{B}_x = B_x, \qquad \bar{B}_y = B_y + \beta E_z, \qquad \bar{B}_z = B_z - \beta E_y; \tag{3.28}$$

$$\frac{\partial}{\partial \bar{x}} = \frac{\partial}{\partial x} + \beta \frac{\partial}{\partial t}, \qquad \frac{\partial}{\partial \bar{y}} = \frac{\partial}{\partial y}, \qquad \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial z}. \tag{3.29}$$

Substitute into the condition of zero divergence in the rocket frame. Recover the original condition of zero divergence in the laboratory frame, plus the following additional information (requirement for the vanishing of the coefficient of the arbitrary small velocity β):

$$\frac{\partial B_x}{\partial t} + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = 0. \tag{3.30}$$

Had the velocity of transformation been directed in the y- or z-directions, a similar equation would have been obtained for $\partial B_y/\partial t$ or $\partial B_z/\partial t$. In the language of three-dimensional vectors, these three equations reduce to the one equation

$$\frac{\partial \boldsymbol{B}}{\partial t} + \nabla \times \boldsymbol{E} \equiv \frac{\partial \boldsymbol{B}}{\partial t} + \operatorname{curl} \boldsymbol{E} = 0. \tag{3.31}$$

How beautiful that (1) the principle of covariance (laws of physics are the same in every Lorentz reference system, which is equivalent to the geometric view of physics) plus (2) the principle that magnetic tubes of force never end, gives (3) Maxwell's dynamic law for the time-rate of change of the magnetic field! This suggests that the magnetostatic law $\nabla \cdot \boldsymbol{B} = 0$ and the magnetodynamic law $\partial \boldsymbol{B}/\partial t + \nabla \times \boldsymbol{E} = 0$ must be wrapped up together in a single frame-independent, geometric law. In terms of components of the field tensor $\boldsymbol{F}$, that geometric law must read

Magnetodynamics and magnetostatics unified in one geometric law

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0, \tag{3.32}$$

since this reduces to $\nabla \cdot \boldsymbol{B} = 0$ when one takes $\alpha = 1, \beta = 2, \gamma = 3$; and it reduces to $\partial \boldsymbol{B}/\partial t + \nabla \times \boldsymbol{E} = 0$ when one sets any index, e.g., α, equal to zero (see exercise 3.7 below). In frame-independent geometric language, this law is written (see §3.5, exercise 3.14, and Chapter 4 for notation)

$$\boldsymbol{dF} = 0, \text{ or, equivalently, } \boldsymbol{\nabla} \cdot {}^*\boldsymbol{F} = 0; \tag{3.33}$$

and it says, "Take the electromagnetic 2-form $\boldsymbol{F}$ (a geometric object defined even in absence of coordinates); from it construct a new geometric object $\boldsymbol{dF}$ (called the "exterior derivative of $\boldsymbol{F}$"); $\boldsymbol{dF}$ must vanish. The details of this coordinate-free process will be spelled out in exercise 3.15 and in §4.5 (track 2).

Two of Maxwell's equations remain: the electrostatic equation

$$\boldsymbol{\nabla} \cdot \boldsymbol{E} = 4\pi\rho, \tag{3.34}$$

and the electrodynamic equation

$$\partial \boldsymbol{E}/\partial t - \boldsymbol{\nabla} \times \boldsymbol{B} = -4\pi \boldsymbol{J}. \tag{3.35}$$

They, like the magnetostatic and magnetodynamic equations, are actually two different parts of a single geometric law. Written in terms of field components, that law says

Electrodynamics and electrostatics unified in one geometric law

$$F^{\alpha\beta}{}_{,\beta} = 4\pi J^\alpha, \tag{3.36}$$

where the components of the "4-current" $\boldsymbol{J}$ are

$$\begin{aligned} J^0 = \rho &= \text{charge density}, \\ (J^1, J^2, J^3) &= \text{components of current density}. \end{aligned} \tag{3.37}$$

Written in coordinate-free, geometric language, this electrodynamic law says

$$\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J} \text{ or, equivalently, } \boldsymbol{\nabla} \cdot \boldsymbol{F} = 4\pi \boldsymbol{J}. \tag{3.38}$$

(For full discussion, see exercise 3.15 and §4.5, which is on Track 2.)

EXERCISE

Exercise 3.7. MAXWELL'S EQUATIONS

Show, by explicit examination of components, that the geometric laws

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0, \qquad F^{\alpha\beta}{}_{,\beta} = 4\pi J^\alpha,$$

do reduce to Maxwell's equations (3.26), (3.31), (3.34), (3.35), as claimed above.

---

§3.5 WORKING WITH TENSORS

Another mathematical digression is needed. Given an arbitrary tensor field, $\boldsymbol{S}$, of arbitrary rank (choose rank = 3 for concreteness), one can construct new tensor fields by a variety of operations.

Ways to produce new tensors from old:

Gradient

One operation is the *gradient* $\boldsymbol{\nabla}$. (The symbol $\boldsymbol{d}$ is reserved for gradients of scalars, in which case $\boldsymbol{\nabla} f \equiv \boldsymbol{d}f$, and for "exterior derivatives of differential forms;" a Track-2 concept, on which see §4.5.) Like $\boldsymbol{S}$, $\boldsymbol{\nabla S}$ is a machine. It has four slots, whereas $\boldsymbol{S}$ has three. It describes how $\boldsymbol{S}$ changes from point to point. Specifically, if one desires to know how the number $\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ for *fixed* $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ changes under a displacement $\boldsymbol{\xi}$, one inserts $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, $\boldsymbol{\xi}$ into the four slots of $\boldsymbol{\nabla S}$:

$$\begin{aligned}\boldsymbol{\nabla S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{\xi}) &\equiv \partial_{\boldsymbol{\xi}} \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ with } \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \text{ fixed} \\ &\simeq + [\text{value of } \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ at tip of } \boldsymbol{\xi}] \\ &\quad - [\text{value of } \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ at tail of } \boldsymbol{\xi}].\end{aligned} \tag{3.39}$$

In component notation in a Lorentz frame, this says

$$\begin{aligned}\boldsymbol{\nabla S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{\xi}) \equiv \partial_{\boldsymbol{\xi}}(S_{\alpha\beta\gamma} u^\alpha v^\beta w^\gamma) &= \left(\frac{\partial S_{\alpha\beta\gamma}}{\partial x^\delta} \xi^\delta\right) u^\alpha v^\beta w^\gamma \\ &= S_{\alpha\beta\gamma,\delta} u^\alpha v^\beta w^\gamma \xi^\delta.\end{aligned}$$

Thus, the Lorentz-frame components of $\boldsymbol{\nabla S}$ are nothing but the partial derivatives of the components of $\boldsymbol{S}$. Notice that the gradient raises the rank of a tensor by 1 (from 3 to 4 for $\boldsymbol{S}$).

Contraction

Contraction is another process that produces a new tensor from an old one. It seals off ("contracts") two of the old tensor's slots, thereby reducing the rank by two. Specifically, if $\boldsymbol{R}$ is a fourth-rank tensor and $\boldsymbol{M}$ is obtained by contracting the first and third slots of $\boldsymbol{R}$, then the output of $\boldsymbol{M}$ is given by (definition!)

$$\boldsymbol{M}(\boldsymbol{u}, \boldsymbol{v}) = \sum_{\alpha=0}^{3} \boldsymbol{R}(\boldsymbol{e}_\alpha, \boldsymbol{u}, \boldsymbol{\omega}^\alpha, \boldsymbol{v}). \tag{3.40}$$

Here $\boldsymbol{e}_\alpha$ and $\boldsymbol{\omega}^\alpha$ are the basis vectors and 1-forms of a specific but arbitrary Lorentz coordinate frame. It makes no difference which frame is chosen; the result will always be the same (exercise 3.8 below). In terms of components in any Lorentz frame, equation (3.40) says (exercise 3.8)

$$\boldsymbol{M}(\boldsymbol{u}, \boldsymbol{v}) = M_{\mu\nu} u^\mu v^\nu = R_{\alpha\mu}{}^{\alpha}{}_{\nu} u^\mu v^\nu,$$

so that

$$M_{\mu\nu} = R_{\alpha\mu}{}^{\alpha}{}_{\nu}. \tag{3.41}$$

Thus, in terms of components, contraction amounts to putting one index up and the other down, and then summing on them.

Divergence

Divergence is a third process for creating new tensors from old. It is accomplished by taking the gradient, then contracting the gradient's slot with one of the original slots:

$$(\text{divergence of } \boldsymbol{S} \text{ on first slot}) \equiv \boldsymbol{\nabla} \cdot \boldsymbol{S} \qquad \text{is a machine such that}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{\nabla S}(\boldsymbol{\omega}^\alpha, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{e}_\alpha) = S^\alpha{}_{\beta\gamma,\alpha} u^\beta v^\gamma; \tag{3.42}$$

i.e. $\boldsymbol{\nabla} \cdot \boldsymbol{S}$ has components $S^\alpha{}_{\beta\gamma,\alpha}$.

Transpose is a fourth, rather trivial process for creating new tensors. It merely interchanges two slots: Transpose

$\boldsymbol{N}$ obtained by transposing second and third slots of $\boldsymbol{S} \Longrightarrow$

$$\boldsymbol{N}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}). \tag{3.43}$$

Symmetrization and *antisymmetrization* are fifth and sixth processes for producing new tensors from old. A tensor is completely symmetric if its output is unaffected by an interchange of two input vectors or 1-forms: Symmetrization and antisymmetrization

$$\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{S}(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}) = \boldsymbol{S}(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u}) = \cdots. \tag{3.44a}$$

It is completely antisymmetric if it reverses sign on each interchange of input

$$\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = -\boldsymbol{S}(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}) = +\boldsymbol{S}(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u}) = \cdots. \tag{3.44b}$$

Any tensor can be symmetrized or antisymmetrized by constructing an appropriate linear combination of it and its transposes; see exercise 3.12.

Wedge product is a seventh process for producing new tensors from old. It is merely an antisymmetrized tensor product: given two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, their wedge product, the *"bivector"* $\boldsymbol{u} \wedge \boldsymbol{v}$, is defined by Wedge product Bivector

$$\boldsymbol{u} \wedge \boldsymbol{v} \equiv \boldsymbol{u} \otimes \boldsymbol{v} - \boldsymbol{v} \otimes \boldsymbol{u}; \tag{3.45a}$$

similarly, the *"2-form"* $\boldsymbol{\alpha} \wedge \boldsymbol{\beta}$ constructed from two 1-forms is 2-form

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \equiv \boldsymbol{\alpha} \otimes \boldsymbol{\beta} - \boldsymbol{\beta} \otimes \boldsymbol{\alpha}. \tag{3.45b}$$

From three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ one constructs the *"trivector"* Trivector

$$\begin{aligned}\boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w} \equiv (\boldsymbol{u} \wedge \boldsymbol{v}) \wedge \boldsymbol{w} \equiv \boldsymbol{u} \wedge (\boldsymbol{v} \wedge \boldsymbol{w}) \\ &= \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} + \text{terms that guarantee complete antisymmetry} \\ &= \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} + \boldsymbol{v} \otimes \boldsymbol{w} \otimes \boldsymbol{u} + \boldsymbol{w} \otimes \boldsymbol{u} \otimes \boldsymbol{v} \\ &\quad - \boldsymbol{v} \otimes \boldsymbol{u} \otimes \boldsymbol{w} - \boldsymbol{u} \otimes \boldsymbol{w} \otimes \boldsymbol{v} - \boldsymbol{w} \otimes \boldsymbol{v} \otimes \boldsymbol{u}.\end{aligned} \tag{3.45c}$$

From 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ one similarly constructs the "3-forms" $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \boldsymbol{\gamma}$. The wedge product gives a simple way to test for coplanarity (linear dependence) of vectors: if $\boldsymbol{u}$ and $\boldsymbol{v}$ are collinear, so $\boldsymbol{u} = a\boldsymbol{v}$, then

$$\boldsymbol{u} \wedge \boldsymbol{v} = a\boldsymbol{v} \wedge \boldsymbol{v} = 0 \qquad \text{(by antisymmetry of "}\wedge\text{")}.$$

If $\boldsymbol{w}$ is coplanar with $\boldsymbol{u}$ and $\boldsymbol{v}$ so $\boldsymbol{w} = a\boldsymbol{u} + b\boldsymbol{v}$ ("collapsed box"), then

$$\boldsymbol{w} \wedge \boldsymbol{u} \wedge \boldsymbol{v} = a\boldsymbol{u} \wedge \boldsymbol{u} \wedge \boldsymbol{v} + b\boldsymbol{v} \wedge \boldsymbol{u} \wedge \boldsymbol{v} = 0.$$

The symbol "$\wedge$" is called a "hat" or "wedge" or "exterior product sign." Its properties are investigated in Chapter 4.

Taking the dual is an eighth process for constructing new tensors. It plays a fundamental role in Track 2 of this book, but since it is not needed for Track 1, its definition and properties are treated only in the exercises (3.14 and 3.15). Dual

Because the frame-independent geometric notation is somewhat ambiguous (which slots are being contracted? on which slot is the divergence taken? which slots are being transposed?), one often uses component notation to express coordinate-independent, geometric relations between geometric objects. For example,

$$J_{\beta\gamma} = S^{\alpha}{}_{\beta\gamma,\alpha}$$

means "$\boldsymbol{J}$ is a tensor obtained by taking the divergence on the first slot of the tensor $\boldsymbol{S}$". Also,

$$v^{\gamma} = (F_{\mu\nu}F^{\mu\nu})^{,\gamma} \equiv (F_{\mu\nu}F^{\mu\nu})_{,\beta}\eta^{\beta\gamma}$$

means "$\boldsymbol{v}$ is a vector obtained by (1) constructing the tensor product $\boldsymbol{F} \otimes \boldsymbol{F}$ of $\boldsymbol{F}$ with itself, (2) contracting $\boldsymbol{F} \otimes \boldsymbol{F}$ on its first and third slots, and also on its second and fourth, (3) taking the gradient of the resultant scalar function, (4) converting that gradient, which is a 1-form, into the corresponding vector."

Index gymnastics

"Index gymnastics," the technique of extracting the content from geometric equations by working in component notation and rearranging indices as required, must be mastered if one wishes to do difficult calculations in relativity, special or general. Box 3.3 expounds some of the short cuts in index gymnastics, and exercises 3.8–3.18 offer practice.

EXERCISES

Exercise 3.8. CONTRACTION IS FRAME-INDEPENDENT

Show that contraction, as defined in equation (3.40), does not depend on which Lorentz frame $\boldsymbol{e}_{\alpha}$ and $\boldsymbol{\omega}^{\alpha}$ are taken from. Also show that equation (3.40) implies

$$\boldsymbol{M}(\boldsymbol{u}, \boldsymbol{v}) = R_{\alpha\mu}{}^{\alpha}{}_{\nu}u^{\mu}v^{\nu}.$$

Exercise 3.9. DIFFERENTIATION

(a) Justify the formula

$$d(u_{\mu}v^{\nu})/d\tau = (du_{\mu}/d\tau)v^{\nu} + u_{\mu}(dv^{\nu}/d\tau).$$

by considering the special case $\mu = 0$, $\nu = 1$.

(b) Explain why

$$(T^{\alpha\beta}v_{\beta})_{,\mu} = T^{\alpha\beta}{}_{,\mu}v_{\beta} + T^{\alpha\beta}v_{\beta,\mu}.$$

Exercise 3.10. MORE DIFFERENTIATION

(a) Justify the formula,

$$d(u^{\mu}u_{\mu})/d\tau = 2u_{\mu}(du^{\mu}/d\tau),$$

by writing out the summation $u^{\mu}u_{\mu} \equiv \eta_{\mu\nu}u^{\mu}u^{\nu}$ explicitly.

(b) Let δ indicate a variation or small change, and justify the formula

$$\delta(F_{\alpha\beta}F^{\alpha\beta}) = 2F_{\alpha\beta}\delta F^{\alpha\beta}.$$

(c) Compute $(F_{\alpha\beta}F^{\alpha\beta})_{,\mu} = ?$

Box 3.3 TECHNIQUES OF INDEX GYMNASTICS

Equation	*Name and Discussion*
$S^{\alpha}{}_{\beta\gamma} = S(\boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta}, \boldsymbol{e}_{\gamma})$	Computing components.
$S^{\alpha\beta}{}_{\gamma} = S(\boldsymbol{\omega}^{\alpha}, \boldsymbol{\omega}^{\beta}, \boldsymbol{e}_{\gamma})$	Computing other components.
$\boldsymbol{S} = S^{\alpha}{}_{\beta\gamma}\boldsymbol{e}_{\alpha} \otimes \boldsymbol{\omega}^{\beta} \otimes \boldsymbol{\omega}^{\gamma}$	Reconstructing the rank-$\binom{1}{2}$ version of $\boldsymbol{S}$.
$\boldsymbol{S} = S^{\alpha\beta\gamma}\boldsymbol{e}_{\alpha} \otimes \boldsymbol{e}_{\beta} \otimes \boldsymbol{e}_{\gamma}$	Reconstructing the rank-$\binom{3}{0}$ version of $\boldsymbol{S}$. [Recall: one does not usually distinguish between the various versions; see equation (3.15) and associated discussion.]
$S^{\alpha\beta}{}_{\gamma} = \eta^{\beta\mu}S^{\alpha}{}_{\mu\gamma}$	Raising an index.
$S^{\alpha}{}_{\mu\gamma} = \eta_{\mu\beta}S^{\alpha\beta}{}_{\gamma}$	Lowering an index.
$M_{\mu} = S^{\alpha}{}_{\mu\alpha}$	Contraction of $\boldsymbol{S}$ to form a new tensor $\boldsymbol{M}$.
$T^{\alpha\beta}{}_{\mu\nu} = S^{\alpha\beta}{}_{\mu}M_{\nu}$	Tensor product of $\boldsymbol{S}$ with $\boldsymbol{M}$ to form a new tensor $\boldsymbol{T}$.
$\boldsymbol{A}^2 = A^{\alpha}A_{\alpha}$	Squared length of vector $\boldsymbol{A}$ produced by forming tensor product $\boldsymbol{A} \otimes \boldsymbol{A}$ and then contracting, which is the same as forming the corresponding 1-form $\tilde{\boldsymbol{A}}$ and then piercing: $\boldsymbol{A}^2 = \langle \tilde{\boldsymbol{A}}, \boldsymbol{A} \rangle = A^{\alpha}A_{\alpha}$.
$\eta_{\alpha\beta}\eta^{\beta\gamma} = \delta_{\alpha}{}^{\gamma}$	The matrix formed from the metric's "covariant components," $\lVert\eta_{\alpha\beta}\rVert$, is the inverse of that formed from its "contravariant components," $\lVert\eta^{\alpha\beta}\rVert$. Equivalently, raising one index of the metric $\eta_{\alpha\beta}$ produces the Kronecker delta.
$S^{\alpha}{}_{\beta\gamma} = N^{\alpha}{}_{\beta,\gamma}$	Gradient of $\boldsymbol{N}$ to form a new tensor $\boldsymbol{S}$.
$R_{\beta} = N^{\alpha}{}_{\beta,\alpha}$	Divergence of $\boldsymbol{N}$ to form a new tensor $\boldsymbol{R}$.
$N^{\alpha}{}_{\beta,\gamma} = (\eta_{\beta\mu}N^{\alpha\mu})_{,\gamma} = \eta_{\beta\mu}N^{\alpha\mu}{}_{,\gamma}$	Taking gradients and raising or lowering indices are operations that commute.
$N^{\alpha}{}_{\beta}{}^{,\gamma} \equiv N^{\alpha}{}_{\beta,\mu}\eta^{\mu\gamma}$	Contravariant index on a gradient is obtained by raising covariant index.
$(R_{\alpha}M_{\beta})_{,\gamma} = R_{\alpha,\gamma}M_{\beta} + R_{\alpha}M_{\beta,\gamma}$	Gradient of a tensor product; says $\boldsymbol{\nabla}(\boldsymbol{R} \otimes \boldsymbol{M}) =$ Transpose $(\boldsymbol{\nabla}\boldsymbol{R} \otimes \boldsymbol{M}) + \boldsymbol{R} \otimes \boldsymbol{\nabla}\boldsymbol{M}$.
$G_{\alpha\beta} = F_{[\alpha\beta]} \equiv \frac{1}{2}(F_{\alpha\beta} - F_{\beta\alpha})$	Antisymmetrizing a tensor $\boldsymbol{F}$ to produce a new tensor $\boldsymbol{G}$.
$H_{\alpha\beta} = F_{(\alpha\beta)} \equiv \frac{1}{2}(F_{\alpha\beta} + F_{\beta\alpha})$	Symmetrizing a tensor $\boldsymbol{F}$ to produce a new tensor $\boldsymbol{H}$.
${}^{*}J_{\alpha\beta\gamma} = J^{\mu}\varepsilon_{\mu\alpha\beta\gamma}$	Forming the rank-3 tensor that is dual to a vector (exercise 3.14).
${}^{*}F_{\alpha\beta} = \frac{1}{2}F^{\mu\nu}\varepsilon_{\mu\nu\alpha\beta}$	Forming the antisymmetric rank-2 tensor that is dual to a given antisymmetric rank-2 tensor (exercise 3.14).
${}^{*}B_{\alpha} = \frac{1}{6}B^{\lambda\mu\nu}\varepsilon_{\lambda\mu\nu\alpha}$	Forming the 1-form that is dual to an antisymmetric rank-3 tensor (exercise 3.14).

Exercise 3.11. SYMMETRIES

Let $A_{\mu\nu}$ be an antisymmetric tensor so that $A_{\mu\nu} = -A_{\nu\mu}$; and let $S^{\mu\nu}$ be a symmetric tensor so that $S^{\mu\nu} = S^{\nu\mu}$.

(a) Justify the equations $A_{\mu\nu}S^{\mu\nu} = 0$ in two ways: first, by writing out the sum explicitly (all sixteen terms) and showing how the terms in the sum cancel in pairs; second, by giving an argument to justify each equals sign in the following string:

$$A_{\mu\nu}S^{\mu\nu} = -A_{\nu\mu}S^{\mu\nu} = -A_{\nu\mu}S^{\nu\mu} = -A_{\alpha\beta}S^{\alpha\beta} = -A_{\mu\nu}S^{\mu\nu} = 0.$$

(b) Establish the following two identities for any arbitrary tensor $V_{\mu\nu}$:

$$V^{\mu\nu}A_{\mu\nu} = \frac{1}{2}(V^{\mu\nu} - V^{\nu\mu})A_{\mu\nu}, \qquad V^{\mu\nu}S_{\mu\nu} = \frac{1}{2}(V^{\mu\nu} + V^{\nu\mu})S_{\mu\nu}.$$

Exercise 3.12. SYMMETRIZATION AND ANTISYMMETRIZATION

To "symmetrize" a tensor, one averages it with all of its transposes. The components of the new, symmetrized tensor are distinguished by round brackets:

$$\begin{aligned} V_{(\mu\nu)} &\equiv \frac{1}{2}(V_{\mu\nu} + V_{\nu\mu}); \\ V_{(\mu\nu\lambda)} &\equiv \frac{1}{3!}(V_{\mu\nu\lambda} + V_{\nu\lambda\mu} + V_{\lambda\mu\nu} + V_{\nu\mu\lambda} + V_{\mu\lambda\nu} + V_{\lambda\nu\mu}). \end{aligned} \tag{3.46}$$

One "antisymmetrizes" a tensor (square brackets) similarly:

$$\begin{aligned} V_{[\mu\nu]} &\equiv \frac{1}{2}(V_{\mu\nu} - V_{\nu\mu}); \\ V_{[\mu\nu\lambda]} &\equiv \frac{1}{3!}(V_{\mu\nu\lambda} + V_{\nu\lambda\mu} + V_{\lambda\mu\nu} - V_{\nu\mu\lambda} - V_{\mu\lambda\nu} - V_{\lambda\nu\mu}). \end{aligned} \tag{3.47}$$

(a) Show that such symmetrized and antisymmetrized tensors are, indeed, symmetric and antisymmetric under interchange of the vectors inserted into their slots:

$$\begin{aligned} V_{(\alpha\beta\gamma)}u^\alpha v^\beta w^\gamma &= +V_{(\alpha\beta\gamma)}v^\alpha u^\beta w^\gamma = \cdots, \\ V_{[\alpha\beta\gamma]}u^\alpha v^\beta w^\gamma &= -V_{[\alpha\beta\gamma]}v^\alpha u^\beta w^\gamma = \cdots. \end{aligned}$$

(b) Show that a second-rank tensor can be reconstructed from its symmetric and antisymmetric parts,

$$V_{\mu\nu} = V_{(\mu\nu)} + V_{[\mu\nu]}, \tag{3.48}$$

but that a third-rank tensor cannot; $V_{(\alpha\beta\gamma)}$ and $V_{[\alpha\beta\gamma]}$ contain together "less information" than $V_{\alpha\beta\gamma}$. "Young diagrams" (see, e.g., Messiah [1961], appendix D) describe other symmetries, more subtle than these two, which contain the missing information.

(c) Show that the electromagnetic field tensor satisfies

$$F_{(\alpha\beta)} = 0, \qquad F_{\alpha\beta} = F_{[\alpha\beta]}. \tag{3.49a}$$

(d) Show that Maxwell's "magnetic" equations

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0$$

can be rewritten in the form

$$F_{[\alpha\beta,\gamma]} = 0. \tag{3.49b}$$

Exercise 3.13. LEVI-CIVITA TENSOR

The "Levi-Civita tensor" $\boldsymbol{\varepsilon}$ in spacetime is a fourth-rank, completely antisymmetric tensor:

$$\boldsymbol{\varepsilon}(\boldsymbol{n}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ changes sign when any two of the vectors are interchanged.} \tag{3.50a}$$

Choose an arbitrary but specific Lorentz frame, with $\boldsymbol{e}_0$ pointing toward the future and with $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ a righthanded set of spatial basis vectors. The covariant components of $\boldsymbol{\varepsilon}$ in this frame are

$$\varepsilon_{0123} = \boldsymbol{\varepsilon}(\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3) = +1. \tag{3.50b}$$

[Note: In an n-dimensional space, $\boldsymbol{\varepsilon}$ is the analogous completely antisymmetric rank-n tensor. Its components are

$$\varepsilon_{12\ldots n} = \boldsymbol{\varepsilon}(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n) = +1, \tag{3.50c}$$

when computed on a "positively oriented," orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.]

(a) Use the antisymmetry to show that

$$\varepsilon_{\alpha\beta\gamma\delta} = 0 \text{ unless } \alpha, \beta, \gamma, \delta \text{ are all different,} \tag{3.50d}$$

$$\varepsilon_{\pi_0\pi_1\pi_2\pi_3} = \begin{cases} +1 \text{ for even permutations of } 0, 1, 2, 3, \text{ and} \\ -1 \text{ for odd permutations.} \end{cases} \tag{3.50e}$$

(b) Show that

$$\varepsilon^{\pi_0\pi_1\pi_2\pi_3} = -\varepsilon_{\pi_0\pi_1\pi_2\pi_3}. \tag{3.50f}$$

(c) By means of a Lorentz transformation show that $\varepsilon^{\bar{\alpha}\bar{\beta}\bar{\gamma}\bar{\delta}}$ and $\varepsilon_{\bar{\alpha}\bar{\beta}\bar{\gamma}\bar{\delta}}$ have these same values in any other Lorentz frame with $\boldsymbol{e}_{\bar{0}}$ pointing toward the future and with $\boldsymbol{e}_{\bar{1}}$, $\boldsymbol{e}_{\bar{2}}$, $\boldsymbol{e}_{\bar{3}}$ a righthanded set. Hint: show that

$$\varepsilon^{\alpha\beta\gamma\delta} \Lambda^{\bar{0}}{}_{\alpha} \Lambda^{\bar{1}}{}_{\beta} \Lambda^{\bar{2}}{}_{\gamma} \Lambda^{\bar{3}}{}_{\delta} = -\det|\Lambda^{\bar{\mu}}{}_{\nu}|; \tag{3.50g}$$

from $\Lambda^T \eta \Lambda = \eta$, show that $\det|\Lambda^{\bar{\mu}}{}_{\nu}| = \pm 1$; and verify that the determinant is $+1$ for transformations between frames with $\boldsymbol{e}_0$ and $\boldsymbol{e}_{\bar{0}}$ future-pointing, and with $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ and $\boldsymbol{e}_{\bar{1}}$, $\boldsymbol{e}_{\bar{2}}$, $\boldsymbol{e}_{\bar{3}}$ righthanded.

(d) What are the components of $\boldsymbol{\varepsilon}$ in a Lorentz frame with past-pointing $\boldsymbol{e}_{\bar{0}}$? with lefthanded $\boldsymbol{e}_{\bar{1}}$, $\boldsymbol{e}_{\bar{2}}$, $\boldsymbol{e}_{\bar{3}}$?

(e) From the Levi-Civita tensor, one can construct several "permutation tensors." In index notation:

$$\delta^{\alpha\beta\gamma}{}_{\mu\nu\lambda} \equiv -\varepsilon^{\alpha\beta\gamma\rho}\varepsilon_{\mu\nu\lambda\rho}; \tag{3.50h}$$

$$\delta^{\alpha\beta}{}_{\mu\nu} \equiv \frac{1}{2}\,\delta^{\alpha\beta\lambda}{}_{\mu\nu\lambda} = -\frac{1}{2}\varepsilon^{\alpha\beta\lambda\rho}\varepsilon_{\mu\nu\lambda\rho}; \tag{3.50i}$$

$$\delta^{\alpha}{}_{\mu} \equiv \frac{1}{3}\,\delta^{\alpha\beta}{}_{\mu\beta} = \frac{1}{6}\,\delta^{\alpha\beta\lambda}{}_{\mu\beta\lambda} = -\frac{1}{6}\,\varepsilon^{\alpha\beta\lambda\rho}\varepsilon_{\mu\beta\lambda\rho}. \tag{3.50j}$$

Show that:

$$\delta^{\alpha\beta\gamma}{}_{\mu\nu\lambda} = \begin{cases} +1 \text{ if } \alpha\beta\gamma \text{ is an even permutation of } \mu\nu\lambda, \\ -1 \text{ if } \alpha\beta\gamma \text{ is an odd permutation of } \mu\nu\lambda, \\ 0 \text{ otherwise;} \end{cases} \tag{3.50k}$$

$$\delta^{\alpha\beta}{}_{\mu\nu} = \delta^{\alpha}{}_{\mu}\delta^{\beta}{}_{\nu} - \delta^{\alpha}{}_{\nu}\delta^{\beta}{}_{\mu}$$

$$= \begin{cases} +1 \text{ if } \alpha\beta \text{ is an even permutation of } \mu\nu, \\ -1 \text{ if } \alpha\beta \text{ is an odd permutation of } \mu\nu, \\ \;\;0 \text{ otherwise;} \end{cases} \tag{3.50l}$$

$$\delta^{\alpha}{}_{\mu} = \begin{cases} +1 \text{ if } \alpha = \mu, \\ \;\;0 \text{ otherwise.} \end{cases} \tag{3.50m}$$

Exercise 3.14. DUALS

From any vector $\boldsymbol{J}$, any second-rank antisymmetric tensor $\boldsymbol{F}(F_{\alpha\beta} = F_{[\alpha\beta]})$, and any third-rank antisymmetric tensor $\boldsymbol{B}(B_{\alpha\beta\gamma} = B_{[\alpha\beta\gamma]})$, one can construct new tensors defined by

$$^{*}J_{\alpha\beta\gamma} = J^{\mu}\varepsilon_{\mu\alpha\beta\gamma}, \qquad ^{*}F_{\alpha\beta} = \frac{1}{2}F^{\mu\nu}\varepsilon_{\mu\nu\alpha\beta}, \qquad ^{*}B_{\alpha} = \frac{1}{3!}B^{\lambda\mu\nu}\varepsilon_{\lambda\mu\nu\alpha}. \tag{3.51}$$

One calls $^{*}\boldsymbol{J}$ the *"dual"* of $\boldsymbol{J}$, $^{*}\boldsymbol{F}$ the dual of $\boldsymbol{F}$, and $^{*}\boldsymbol{B}$ the dual of $\boldsymbol{B}$. [A previous and entirely distinct use of the word "dual" (§2.7) called a *set* of basis one-forms $\{\boldsymbol{\omega}^{\alpha}\}$ dual to a *set* of basis vectors $\{\boldsymbol{e}_{\alpha}\}$ if $\langle\boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta}\rangle = \delta^{\alpha}{}_{\beta}$. Fortunately there are no grounds for confusion between the two types of duality. One relates sets of vectors to sets of 1-forms. The other relates antisymmetric tensors of rank p to antisymmetric tensors of rank $4 - p$.]

(a) Show that

$$^{**}\boldsymbol{J} = \boldsymbol{J}, \qquad ^{**}\boldsymbol{F} = -\boldsymbol{F}, \qquad ^{**}\boldsymbol{B} = \boldsymbol{B}. \tag{3.52}$$

so (aside from sign) one can recover any completely antisymmetric tensor $\boldsymbol{H}$ from its dual $^{*}\boldsymbol{H}$ by taking the dual once again, $^{**}\boldsymbol{H}$. This shows that $\boldsymbol{H}$ and $^{*}\boldsymbol{H}$ contain precisely the same information.

(b) Make explicit this fact of same-information-content by writing out the components $^{*}A^{\alpha\beta\gamma}$ in terms of A^{α}, also $^{*}F^{\alpha\beta}$ in terms of $F^{\alpha\beta}$, also $^{*}B^{\alpha}$ in terms of $B^{\alpha\beta\gamma}$.

Exercise 3.15. GEOMETRIC VERSIONS OF MAXWELL EQUATIONS

Show that, if $\boldsymbol{F}$ is the electromagnetic field tensor, then $\boldsymbol{\nabla} \cdot {}^{*}\boldsymbol{F} = 0$ is a geometric frame-independent version of the Maxwell equations

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0.$$

Similarly show that $\boldsymbol{\nabla} \cdot \boldsymbol{F} = 4\pi\boldsymbol{J}$ (divergence on second slot of $\boldsymbol{F}$) is a geometric version of $F^{\alpha\beta}{}_{,\beta} = 4\pi J^{\alpha}$.

Exercise 3.16. CHARGE CONSERVATION

From Maxwell's equations $F^{\alpha\beta}{}_{,\beta} = 4\pi J^{\alpha}$, derive the "equation of charge conservation"

$$J^{\alpha}{}_{,\alpha} = 0. \tag{3.53}$$

Show that this equation does, indeed, correspond to conservation of charge. It will be studied further in Chapter 5.

Exercise 3.17. VECTOR POTENTIAL

The vector potential $\boldsymbol{A}$ of electromagnetic theory generates the electromagnetic field tensor via the geometric equation

$$\boldsymbol{F} = -(\text{antisymmetric part of } \boldsymbol{\nabla A}), \tag{3.54}$$

i.e.,

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}. \tag{3.54'}$$

(a) Show that the electric and magnetic fields in a specific Lorentz frame are given by

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}, \qquad \boldsymbol{E} = -\partial \boldsymbol{A}/\partial t - \nabla A^0. \tag{3.55}$$

(b) Show that $\boldsymbol{F}$ will satisfy Maxwell's equations if and only if $\boldsymbol{A}$ satisfies

$$A^{\alpha,\mu}{}_{,\mu} - A^{\mu}{}_{,\mu}{}^{,\alpha} = -4\pi J^{\alpha}. \tag{3.56}$$

(c) Show that "gauge transformations"

$$\boldsymbol{A}_{\text{NEW}} = \boldsymbol{A}_{\text{OLD}} + \boldsymbol{d}\phi, \qquad \phi = \text{arbitrary function}, \tag{3.57}$$

leave $\boldsymbol{F}$ unaffected.

(d) Show that one can adjust the gauge so that

$$\nabla \cdot \boldsymbol{A} = 0 \qquad \text{("Lorentz gauge")}, \tag{3.58a}$$

$$\Box \boldsymbol{A} = -4\pi \boldsymbol{J}. \tag{3.58b}$$

Here $\Box$ is the wave operator ("d'Alembertian"):

$$\Box \boldsymbol{A} = A^{\alpha,\mu}{}_{,\mu}\boldsymbol{e}_{\alpha}. \tag{3.59}$$

Exercise 3.18. DIVERGENCE OF ELECTROMAGNETIC STRESS-ENERGY TENSOR

From an electromagnetic field tensor $\boldsymbol{F}$, one constructs a second-rank, symmetric tensor $\boldsymbol{T}$ (**"stress**-energy tensor," to be studied in Chapter 5) as follows:

$$T^{\mu\nu} = \frac{1}{4\pi}\left(F^{\mu\alpha}F^{\nu}{}_{\alpha} - \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right). \tag{3.60}$$

As an exercise in index gymnastics:

(a) Show that $\nabla \cdot \boldsymbol{T}$ has components

$$T^{\mu\nu}{}_{,\nu} = \frac{1}{4\pi}\left[F^{\mu\alpha}{}_{,\nu}F^{\nu}{}_{\alpha} + F^{\mu\alpha}F^{\nu}{}_{\alpha,\nu} - \frac{1}{2}F_{\alpha\beta}{}^{,\mu}F^{\alpha\beta}\right]. \tag{3.61}$$

(b) Manipulate this expression into the form

$$T_{\mu}{}^{\nu}{}_{,\nu} = \frac{1}{4\pi}\left[-F_{\mu\alpha}F^{\alpha\nu}{}_{,\nu} - \frac{1}{2}F^{\alpha\beta}(F_{\alpha\beta,\mu} + F_{\mu\alpha,\beta} + F_{\beta\mu,\alpha})\right]; \tag{3.62}$$

note that the first term of (3.62) arises directly from the second term of (3.61).

(c) Use Maxwell's equations to conclude that

$$T^{\mu\nu}{}_{,\nu} = -F^{\mu\alpha}J_{\alpha}. \tag{3.63}$$

2

CHAPTER 4

ELECTROMAGNETISM AND DIFFERENTIAL FORMS

The ether trembled at his agitations
In a manner so familiar that I only need to say,
In accordance with Clerk Maxwell's six equations
It tickled peoples' optics far away.
You can feel the way it's done,
You may trace them as they run—
$d\gamma$ *by* dy *less* $d\beta$ *by* dz *is equal* KdX/dt . . .

While the curl of (X, Y, Z) *is the*
minus d/dt *of the vector* (a, b, c).

From *The Revolution of the Corpuscle,*
written by A. A. Robb
(to the tune of *The Interfering Parrott*)
for a dinner of the research students
of the Cavendish Laboratory
in the days of the old mathematics.

This chapter is all Track 2. It is needed as preparation for §§14.5 and 14.6 (computation of curvature using differential forms) and for Chapter 15 (Bianchi identities and boundary of a boundary), but is not needed for the rest of the book.

§4.1. EXTERIOR CALCULUS

Stacks of surfaces, individually or intersecting to make "honeycombs," "egg crates," and other such structures ("differential forms"), give unique insight into the geometry of electromagnetism and gravitation. However, such insight comes at some cost in time. Therefore, most readers should skip this chapter and later material that depends on it during a first reading of this book.

Analytically speaking, differential forms are completely antisymmetric tensors; pictorially speaking, they are intersecting stacks of surfaces. The mathematical formalism for manipulating differential forms with ease, called "exterior calculus," is summarized concisely in Box 4.1; its basic features are illustrated in the rest of this chapter by rewriting electromagnetic theory in its language. An effective way to tackle this chapter might be to (1) scan Box 4.1 to get the flavor of the formalism; (2) read the rest of the chapter in detail; (3) restudy Box 4.1 carefully; (4) get practice in manipulating the formalism by working the exercises.*

(continued on page 99)

*Exterior calculus is treated in greater detail than here by: É. Cartan (1945); de Rham (1955); Nickerson, Spencer, and Steenrod (1959); Hauser (1970); Israel (1970); especially Flanders (1963, relatively easy, with many applications); Spivak (1965, sophomore or junior level, but fully in tune with modern mathematics); H. Cartan (1970); and Choquet-Bruhat (1968a).

Box 4.1 DIFFERENTIAL FORMS AND EXTERIOR CALCULUS IN BRIEF

The fundamental definitions and formulas of exterior calculus are summarized here for ready reference. Each item consists of a general statement (at left of page) plus a leading application (at right of page). This formalism is applicable not only to spacetime, but also to more general geometrical systems (see heading of each section). No attempt is made here to demonstrate the internal consistency of the formalism, nor to derive it from any set of definitions and axioms. For a systematic treatment that does so, see, e.g., Spivak (1965), or Misner and Wheeler (1957).

A. Algebra I (applicable to any vector space)

1. *Basis 1-forms.*
   a. Coordinate basis $\boldsymbol{\omega}^j = \boldsymbol{d}x^j$
   (j tells which 1-form, not which component).
   b. General basis $\boldsymbol{\omega}^j = L^j{}_{k'}\, \boldsymbol{d}x^{k'}$.

An application
Simple basis 1-forms for analyzing Schwarzschild geometry around static spherically symmetric center of attraction:

$$\boldsymbol{\omega}^0 = (1 - 2m/r)^{1/2}\, \boldsymbol{d}t;$$
$$\boldsymbol{\omega}^1 = (1 - 2m/r)^{-1/2}\, \boldsymbol{d}r;$$
$$\boldsymbol{\omega}^2 = r\, \boldsymbol{d}\theta;$$
$$\boldsymbol{\omega}^3 = r \sin\theta\, \boldsymbol{d}\phi.$$

2. *General* p*-form* (*or* p*-vector*) is a completely antisymmetric tensor of rank $\binom{0}{p}$ [or $\binom{p}{0}$]. It can be expanded in terms of wedge products (see §3.5 and exercise 4.12):

$$\boldsymbol{\alpha} = \frac{1}{p!}\alpha_{i_1 i_2 \dots i_p}\boldsymbol{\omega}^{i_1} \wedge \boldsymbol{\omega}^{i_2} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}$$
$$\equiv \alpha_{|i_1 i_2 \dots i_p|}\boldsymbol{\omega}^{i_1} \wedge \boldsymbol{\omega}^{i_2} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}.$$

(Note: Vertical bars around the indices mean summation extends only over $i_1 < i_2 < \cdots < i_p$.)

Two applications
Energy-momentum 1-form is of type $\boldsymbol{\alpha} = \alpha_i \boldsymbol{\omega}^i$ or

$$\boldsymbol{p} = -E\, \boldsymbol{d}t + p_x\, \boldsymbol{d}x + p_y\, \boldsymbol{d}y + p_z\, \boldsymbol{d}z.$$

Faraday is a 2-form of type $\boldsymbol{\beta} = \beta_{|\mu\nu|}\boldsymbol{\omega}^\mu \wedge \boldsymbol{\omega}^\nu$ or in flat spacetime

$$\boldsymbol{F} = -E_x\, \boldsymbol{d}t \wedge \boldsymbol{d}x - E_y\, \boldsymbol{d}t \wedge \boldsymbol{d}y - E_z\, \boldsymbol{d}t \wedge \boldsymbol{d}z$$
$$+ B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z + B_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y$$

Box 4.1 (*continued*)

3. *Wedge product.*
All familiar rules of addition and multiplication hold, such as

$$(a\boldsymbol{\alpha} + b\boldsymbol{\beta}) \wedge \boldsymbol{\gamma} = a\boldsymbol{\alpha} \wedge \boldsymbol{\gamma} + b\boldsymbol{\beta} \wedge \boldsymbol{\gamma},$$
$$(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) \wedge \boldsymbol{\gamma} = \boldsymbol{\alpha} \wedge (\boldsymbol{\beta} \wedge \boldsymbol{\gamma}) \equiv \boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \boldsymbol{\gamma},$$

except for a modified commutation law between a p-form $\boldsymbol{\alpha}$ and a q-form $\boldsymbol{\beta}$:

$$\underset{p}{\boldsymbol{\alpha}} \wedge \underset{q}{\boldsymbol{\beta}} = (-1)^{pq} \underset{q}{\boldsymbol{\beta}} \wedge \underset{p}{\boldsymbol{\alpha}}.$$

Applications to 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}$:

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = -\boldsymbol{\beta} \wedge \boldsymbol{\alpha}, \qquad \boldsymbol{\alpha} \wedge \boldsymbol{\alpha} = 0;$$
$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = (\alpha_j \boldsymbol{\omega}^j) \wedge (\beta_k \boldsymbol{\omega}^k) = \alpha_j \beta_k \boldsymbol{\omega}^j \wedge \boldsymbol{\omega}^k$$
$$= \frac{1}{2}(\alpha_j \beta_k - \beta_j \alpha_k) \boldsymbol{\omega}^j \wedge \boldsymbol{\omega}^k.$$

4. *Contraction of* p-*form on* p-*vector.*

$$\langle \underset{p}{\boldsymbol{\alpha}}, \underset{p}{\boldsymbol{A}} \rangle$$
$$= \alpha_{|i_1 \ldots i_p|} A^{|j_1 \ldots j_p|} \underbrace{\langle \boldsymbol{\omega}^{i_1} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}, \boldsymbol{e}_{j_1} \wedge \cdots \wedge \boldsymbol{e}_{j_p} \rangle}_{[\equiv \delta^{i_1 \ldots i_p}_{j_1 \ldots j_p} \text{ (see exercises 3.13 and 4.12)}]}$$
$$= \alpha_{|i_1 \ldots i_p|} A^{i_1 \ldots i_p}.$$

Four applications

a. Contraction of a particle's energy-momentum 1-form $\boldsymbol{p} = p_\alpha \boldsymbol{\omega}^\alpha$ with 4-velocity $\boldsymbol{u} = u^\alpha \boldsymbol{e}_\alpha$ of observer (a 1-vector):

$$-\langle \boldsymbol{p}, \boldsymbol{u} \rangle = -p_\alpha u^\alpha = \text{energy of particle.}$$

b. Contraction of ***Faraday*** 2-form $\boldsymbol{F}$ with bivector $\delta\mathcal{P} \wedge \Delta\mathcal{P}$ [where $\delta\mathcal{P} = (d\mathcal{P}/d\lambda_1)\Delta\lambda_1$ and $\Delta\mathcal{P} = (d\mathcal{P}/d\lambda_2)\Delta\lambda_2$ are two infinitesimal vectors in a 2-surface $\mathcal{P}(\lambda_1, \lambda_2)$, and the bivector represents the surface element they span] is the magnetic flux $\Phi = \langle \boldsymbol{F}, \delta\mathcal{P} \wedge \Delta\mathcal{P} \rangle$ through that surface element.

c. More generally, a p-dimensional parallelepiped with vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ for legs has an oriented volume described by the "*simple*" p-vector $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \boldsymbol{a}_p$ (oriented because interchange of two legs changes its sign). An egg-crate type of structure with walls made from the hyperplanes of p different 1-forms $\boldsymbol{\sigma}^1$,

$\boldsymbol{\sigma}^2, \ldots, \boldsymbol{\sigma}^p$ is described by the "*simple*" p-form $\boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p$. The number of cells of $\boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p$ sliced through by the infinitesimal p-volume $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \wedge \boldsymbol{a}_p$ is

$$\langle \boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p, \boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \wedge \boldsymbol{a}_p \rangle.$$

d. The Jacobian determinant of a set of p functions $f^k(x^1, \ldots, x^n)$ with respect to p of their arguments is

$$\left\langle \boldsymbol{d}f^1 \wedge \boldsymbol{d}f^2 \wedge \cdots \wedge \boldsymbol{d}f^p, \frac{\partial \mathscr{P}}{\partial x^1} \wedge \frac{\partial \mathscr{P}}{\partial x^2} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial x^p} \right\rangle = \det \left\| \left(\frac{\partial f^k}{\partial x^j} \right) \right\| \equiv \frac{\partial(f^1, f^2, \ldots, f^p)}{\partial(x^1, x^2, \ldots, x^p)}.$$

5. *Simple forms.*
   a. A simple p-form is one that can be written as a wedge product of p 1-forms:

$$\underset{p}{\boldsymbol{\sigma}} = \underbrace{\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma}}_{p \text{ factors}}.$$

   b. A simple p-form $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma}$ is represented by the intersecting families of surfaces of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ (egg-crate structure) plus a sense of circulation (orientation).

Applications:

a. In four dimensions (e.g., spacetime) all 0-forms, 1-forms, 3-forms, and 4-forms are simple. A 2-form $\boldsymbol{F}$ is generally a sum of two simple forms, e.g., $\boldsymbol{F} = -e\, \boldsymbol{d}t \wedge \boldsymbol{d}x + h\, \boldsymbol{d}y \wedge \boldsymbol{d}z$; it is simple if and only if $\boldsymbol{F} \wedge \boldsymbol{F} = 0$.

b. A set of 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ is linearly dependent (one a linear combination of the others) if and only if

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma} = 0 \qquad \text{(egg crate collapsed).}$$

B. Exterior Derivative (applicable to any "differentiable manifold," with or without metric)

1. $\boldsymbol{d}$ produces a $(p+1)$-form $\boldsymbol{d\sigma}$ from a p-form $\boldsymbol{\sigma}$.
2. Effect of $\boldsymbol{d}$ is defined by induction using the

Box 4.1 *(continued)*

(Chapter 2) definition of $\boldsymbol{d}f$, and f a function (0-form), plus

$$\boldsymbol{d}(\underset{p}{\boldsymbol{\alpha}} \wedge \underset{q}{\boldsymbol{\beta}}) = \boldsymbol{d\alpha} \wedge \boldsymbol{\beta} + (-1)^p \boldsymbol{\alpha} \wedge \boldsymbol{d\beta},$$

$$\boldsymbol{d}^2 = \boldsymbol{dd} = 0.$$

Two applications

$$\boldsymbol{d}(\boldsymbol{\alpha} \wedge \boldsymbol{d\beta}) = \boldsymbol{d\alpha} \wedge \boldsymbol{d\beta}.$$

For the p-form $\boldsymbol{\phi}$, with

$$\boldsymbol{\phi} = \phi_{|i_1 \ldots i_p|}\, \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p},$$

one has (alternative and equivalent definition of $\boldsymbol{d\phi}$)

$$\boldsymbol{d\phi} = \boldsymbol{d}\phi_{|i_1 \ldots i_p|} \wedge \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p}.$$

C. Integration (applicable to any "differentiable manifold," with or without metric)

1. *Pictorial interpretation.*
   Text and pictures of Chapter 4 interpret $\int \boldsymbol{\alpha}$ (integral of specified 1-form $\boldsymbol{\alpha}$ along specified curve from specified starting point to specified end point) as "number of $\boldsymbol{\alpha}$-surfaces pierced on that route"; similarly, they interpret $\int \boldsymbol{\phi}$ (integral of specified 2-form $\boldsymbol{\phi}$ over specified bit of surface on which there is an assigned sense of circulation or "orientation") as "number of cells of the honeycomb-like structure $\boldsymbol{\phi}$ cut through by that surface"; similarly for the egg-crate-like structures that represent 3-forms; etc.
2. *Computational rules for integration.*
   To evaluate $\int \boldsymbol{\alpha}$, the integral of a p-form

   $$\boldsymbol{\alpha} = \alpha_{|i_1 \ldots i_p|}(x^1, \ldots, x^n)\, \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p},$$

   over a p-dimensional surface, proceed in two steps.
   a. Substitute a parameterization of the surface,

   $$x^k(\lambda^1, \ldots, \lambda^p)$$

   into $\boldsymbol{\alpha}$, and collect terms in the form

   $$\boldsymbol{\alpha} = a(\lambda^j)\, \boldsymbol{d}\lambda^1 \wedge \cdots \wedge \boldsymbol{d}\lambda^p$$

   (this is $\boldsymbol{\alpha}$ viewed as a p-form in the p-dimensional surface);

b. Integrate

$$\int \boldsymbol{\alpha} = \int a(\lambda^j)\, d\lambda^1\, d\lambda^2 \ldots d\lambda^p$$

using elementary definition of integration.

Example: See equations (4.12) to (4.14).

3. *The differential geometry of integration.*
Calculate $\int \boldsymbol{\alpha}$ for a p-form $\boldsymbol{\alpha}$ as follows.
   a. Choose the p-dimensional surface $\mathcal{S}$ over which to integrate.
   b. Represent $\mathcal{S}$ by a parametrization giving the generic point of the surface as a function of the parameters, $\mathscr{P}(\lambda^1, \lambda^2, \ldots \lambda^p)$. This fixes the orientation. The same function with $\lambda^1 \leftrightarrow \lambda^2$, $\mathscr{P}(\lambda^2, \lambda^1, \ldots, \lambda^p)$, describes a different (i.e., oppositely oriented) surface, $-\mathcal{S}$.
   c. The infinitesimal parallelepiped

$$\left(\frac{\partial \mathscr{P}}{\partial \lambda^1}\Delta\lambda^1\right) \wedge \left(\frac{\partial \mathscr{P}}{\partial \lambda^2}\Delta\lambda^2\right) \wedge \cdots \wedge \left(\frac{\partial \mathscr{P}}{\partial \lambda^p}\Delta\lambda^p\right)$$

is tangent to the surface. The number of cells of $\boldsymbol{\alpha}$ it slices is

$$\left\langle \boldsymbol{\alpha}, \frac{\partial \mathscr{P}}{\partial \lambda^1} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial \lambda^p}\right\rangle \Delta\lambda^1 \ldots \Delta\lambda^p.$$

This number changes sign if two of the vectors $\partial\mathscr{P}/\partial\lambda^k$ are interchanged, as for an oppositely oriented surface.
   d. The above provides an interpretation motivating the definition

$$\int \boldsymbol{\alpha} \equiv \int\!\!\int \cdots \int \left\langle \boldsymbol{\alpha}, \frac{\partial \mathscr{P}}{\partial \lambda^1} \wedge \frac{\partial \mathscr{P}}{\partial \lambda^2} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial \lambda^p}\right\rangle d\lambda^1\, d\lambda^2 \ldots d\lambda^p.$$

This definition is identified with the computational rule of the preceding section (C.2) in exercise 4.9.

An application
Integrate a gradient $\boldsymbol{d}f$ along a curve, $\mathscr{P}(\lambda)$ from $\mathscr{P}(0)$ to $\mathscr{P}(1)$:

$$\int \boldsymbol{d}f = \int_0^1 \langle \boldsymbol{d}f, d\mathscr{P}/d\lambda\rangle\, d\lambda = \int_0^1 (df/d\lambda)\, d\lambda$$
$$= f[\mathscr{P}(1)] - f[\mathscr{P}(0)].$$

   e. Three different uses for symbol "d": *First,* lightface d in explicit derivative expressions such as

Box 4.1 (*continued*)

d/da, or df/da, or $d\mathscr{P}/da$; neither numerator nor denominator alone has any meaning, but only the full string of symbols. *Second,* lightface d inside an integral sign; e.g., $\int f\, da$. This is an instruction to perform integration, and has no meaning whatsoever without an integral sign; "$\int \ldots d \ldots$" lives as an indivisible unit. *Third,* sans-serif $\boldsymbol{d}$; e.g., $\boldsymbol{d}$ alone, or $\boldsymbol{d}f$, or $\boldsymbol{d\alpha}$. This is an exterior derivative, which converts a p-form into a $(p+1)$-form. Sometimes lightface d is used for the same purpose. Hence, d alone, or df, or dx, is always an exterior derivative unless coupled to an $\int$ sign (*second* use), or coupled to a / sign (*first* use).

4. *The generalized Stokes theorem* (see Box 4.6).
   a. Let $\partial\mathscr{V}$ be the closed p-dimensional boundary of a $(p+1)$-dimensional surface $\mathscr{V}$. Let $\boldsymbol{\sigma}$ be a p-form defined throughout $\mathscr{V}$.

   Then

$$\boxed{\int_{\mathscr{V}} \boldsymbol{d\sigma} = \int_{\partial\mathscr{V}} \boldsymbol{\sigma}}$$

   [integral of p-form $\boldsymbol{\sigma}$ over boundary $\partial\mathscr{V}$ equals integral of $(p+1)$-form $\boldsymbol{d\sigma}$ over interior $\mathscr{V}$].

   b. For the sign to come out right, orientations of $\mathscr{V}$ and $\partial\mathscr{V}$ must agree in this sense: choose coordinates $y^0, y^1, \ldots, y^p$ on a portion of $\mathscr{V}$, with y^0 specialized so $y^0 \leq 0$ in $\mathscr{V}$, and $y^0 = 0$ at the boundary $\partial\mathscr{V}$; then the orientation

$$\frac{\partial\mathscr{P}}{\partial y^0} \wedge \frac{\partial\mathscr{P}}{\partial y^1} \wedge \cdots \wedge \frac{\partial\mathscr{P}}{\partial y^p}$$

   for $\mathscr{V}$ demands the orientation

$$\frac{\partial\mathscr{P}}{\partial y^1} \wedge \cdots \wedge \frac{\partial\mathscr{P}}{\partial y^p}$$

   for $\partial\mathscr{V}$.

   c. Note: For a nonorientable surface, such as a Möbius strip, where a consistent and continuous choice of orientation is impossible, more intricate mathematics is required to give a definition of "∂" for which the Stokes theorem holds.

Applications: Includes as special cases all integral theorems for surfaces of arbitrary dimension in spaces of arbitrary dimension, with or without metric, generalizing all versions of theorems of Stokes and Gauss. Examples:

a. $\mathcal{V}$ a curve, $\partial\mathcal{V}$ its endpoints, $\boldsymbol{\sigma} = f$ a 0-form (function):

$$\int_{\mathcal{V}} \boldsymbol{d}f = \int_0^1 (df/d\lambda)\, d\lambda = \int_{\partial\mathcal{V}} f = f(1) - f(0).$$

b. $\mathcal{V}$ a 2-surface in 3-space, $\partial\mathcal{V}$ its closed-curve boundary, $\boldsymbol{v}$ a 1-form; translated into Euclidean vector notation, the two integrals are

$$\int_{\mathcal{V}} \boldsymbol{dv} = \int_{\mathcal{V}} (\nabla \times v) \cdot \boldsymbol{dS}; \int_{\partial\mathcal{V}} \boldsymbol{v} = \int_{\partial\mathcal{V}} v \cdot \boldsymbol{dl}.$$

c. Other applications in §§5.8, 20.2, 20.3, 20.5, and exercises 4.10, 4.11, 5.2, and below.

D. Algebra II (applicable to any vector space with metric)

1. *Norm of a* p*-form.*

$$\|\boldsymbol{\alpha}\|^2 \equiv \alpha_{|i_1 \ldots i_p|}\alpha^{i_1 \ldots i_p}.$$

Two applications: Norm of a 1-form equals its squared length, $\|\boldsymbol{\alpha}\|^2 = \boldsymbol{\alpha} \cdot \boldsymbol{\alpha}$. Norm of electromagnetic 2-form or ***Faraday***: $\|\boldsymbol{F}\|^2 = \boldsymbol{B}^2 - \boldsymbol{E}^2$.

2. *Dual of a* p*-form.*

a. In an n-dimensional space, the dual of a p-form $\boldsymbol{\alpha}$ is the $(n-p)$-form ${}^*\boldsymbol{\alpha}$, with components

$$({}^*\alpha)_{k_1 \ldots k_{n-p}} = \alpha^{|i_1 \ldots i_p|}\epsilon_{i_1 \ldots i_p k_1 \ldots k_{n-p}}.$$

b. Properties of duals:

$$\begin{aligned} {}^{**}\boldsymbol{\alpha} &= (-1)^{p-1}\boldsymbol{\alpha} \text{ in spacetime;} \\ \boldsymbol{\alpha} \wedge {}^*\boldsymbol{\alpha} &= \|\boldsymbol{\alpha}\|^2\boldsymbol{\varepsilon} \text{ in general.} \end{aligned}$$

c. Note: the definition of $\boldsymbol{\varepsilon}$ (exercise 3.13) entails choosing an orientation of the space, i.e., deciding which orthonormal bases (1) are "right-handed" and thus (2) have $\boldsymbol{\varepsilon}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n) = +1$.

Applications

a. For f a 0-form, ${}^*f = f\boldsymbol{\varepsilon}$, and $\int f d(\text{volume}) = \int {}^*f$.

b. Dual of charge-current 1-form $\boldsymbol{J}$ is charge-current 3-form ${}^*\boldsymbol{J}$. The total charge Q in a 3-dimensional hypersurface region $\mathcal{S}$ is

$$Q(\mathcal{S}) = \int_{\mathcal{S}} {}^*\boldsymbol{J}.$$

Box 4.1 (*continued*)

Conservation of charge is stated locally by $\boldsymbol{d}{}^*\boldsymbol{J} = 0$. Stokes' Theorem goes from this differential conservation law to the integral conservation law,

$$0 = \int_{\mathcal{V}} \boldsymbol{d}{}^*\boldsymbol{J} = \int_{\partial\mathcal{V}} {}^*\boldsymbol{J}.$$

This law is of most interest when $\partial\mathcal{V} = \mathcal{S}_2 - \mathcal{S}_1$ consists of the future $\mathcal{S}_2$ and past $\mathcal{S}_1$ boundaries of a spacetime region, in which case it states $Q(\mathcal{S}_2) = Q(\mathcal{S}_1)$; see exercise 5.2.

c. Dual of electromagnetic field tensor $\boldsymbol{F} =$ ***Faraday*** is ${}^*\boldsymbol{F} =$ ***Maxwell***. From the $\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\,{}^*\boldsymbol{J}$ Maxwell equation, find $4\pi Q = 4\pi\int_{\mathcal{S}} {}^*\boldsymbol{J} = \int_{\mathcal{S}} \boldsymbol{d}{}^*\boldsymbol{F} = \int_{\partial\mathcal{S}} {}^*\boldsymbol{F}$.

3. *Simple forms revisited.*
   a. The dual of a simple form is simple.
   b. Egg crate of ${}^*\boldsymbol{\sigma}$ is perpendicular to egg crate of $\boldsymbol{\sigma} = \boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\mu}$ in this sense:
   (1) pick any vector $\boldsymbol{V}$ lying in intersection of surfaces of $\boldsymbol{\sigma}$

   $$(\langle\boldsymbol{\alpha}, \boldsymbol{V}\rangle = \langle\boldsymbol{\beta}, \boldsymbol{V}\rangle = \cdots = \langle\boldsymbol{\mu}, \boldsymbol{V}\rangle = 0);$$

   (2) pick any vector $\boldsymbol{W}$ lying in intersection of surfaces of ${}^*\boldsymbol{\sigma}$;
   (3) then $\boldsymbol{V}$ and $\boldsymbol{W}$ are necessarily perpendicular: $\boldsymbol{V}\cdot\boldsymbol{W} = 0$.

Example: $\boldsymbol{\sigma} = 3\,\boldsymbol{d}t$ is a simple 1-form in spacetime.
a. ${}^*\boldsymbol{\sigma} = -3\,\boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z$ is a simple 3-form.
b. General vector in surfaces of $\boldsymbol{\sigma}$ is

$$\boldsymbol{V} = V^x\boldsymbol{e}_x + V^y\boldsymbol{e}_y + V^z\boldsymbol{e}_z.$$

c. General vector in intersection of surfaces of ${}^*\boldsymbol{\sigma}$ is

$$\boldsymbol{W} = W^t\boldsymbol{e}_t.$$

d. $\boldsymbol{W}\cdot\boldsymbol{V} = 0$.

§4.2. ELECTROMAGNETIC 2-FORM AND LORENTZ FORCE

The electromagnetic field tensor, ***Faraday*** $= \boldsymbol{F}$, is an antisymmetric second-rank tensor (i.e., 2-form). Instead of expanding it in terms of the tensor products of basis 1-forms,

$$\boldsymbol{F} = F_{\alpha\beta}\, \boldsymbol{d}x^{\alpha} \otimes \boldsymbol{d}x^{\beta},$$

the exterior calculus prefers to expand in terms of antisymmetrized tensor products (*"exterior products,"* exercise 4.1):

Electromagnetic 2-form expressed in terms of exterior products

$$\boldsymbol{F} = \frac{1}{2} F_{\alpha\beta}\, \boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta}, \tag{4.1}$$

$$\boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta} \equiv \boldsymbol{d}x^{\alpha} \otimes \boldsymbol{d}x^{\beta} - \boldsymbol{d}x^{\beta} \otimes \boldsymbol{d}x^{\alpha}. \tag{4.2}$$

Any 2-form (antisymmetric, second-rank tensor) can be so expanded. The symbol "$\wedge$" is variously called a "wedge," a "hat," or an "exterior product sign"; and $\boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta}$ are the "basis 2-forms" of a given Lorentz frame (see §3.5, exercise 3.12, and Box 4.1).

There is no simpler way to illustrate this 2-form representation of the electromagnetic field than to consider a magnetic field in the x-direction:

$$\begin{aligned} F_{yz} &= -F_{zy} = B_x, \\ \boldsymbol{F} &= B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z. \end{aligned} \tag{4.3}$$

The 1-form $\boldsymbol{d}y = \text{grad}\, y$ is the set of surfaces (actually hypersurfaces) $y = 18$ (all t, x, z), $y = 19$ (all t, x, z), $y = 20$ (all t, x, z), etc.; and surfaces uniformly interpolated between them. Similarly for the 1-form $\boldsymbol{d}z$. The intersection between these two sets of surfaces produces a honeycomb-like structure. That structure becomes a "2-form" when it is supplemented by instructions (see arrows in Figure 4.1) that give a "sense of circulation" to each tube of the honeycomb (order of factors in the "wedge product" of equation 4.2; $\boldsymbol{d}y \wedge \boldsymbol{d}z = -\boldsymbol{d}z \wedge \boldsymbol{d}y$). The 2-form $\boldsymbol{F}$ in the example differs from this "basis 2-form" $\boldsymbol{d}y \wedge \boldsymbol{d}z$ only in this respect, that where $\boldsymbol{d}y \wedge \boldsymbol{d}z$ had one tube, the field 2-form has B_x tubes.

A 2-form as a honeycomb of tubes with a sense of circulation

When one considers a tubular structure that twists and turns on its way through spacetime, one must have more components to describe it. The 2-form for the general electromagnetic field can be written as

$$\begin{aligned} \boldsymbol{F} = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + E_y\, \boldsymbol{d}y \wedge \boldsymbol{d}t + E_z\, \boldsymbol{d}z \wedge \boldsymbol{d}t + B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z \\ + B_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y \end{aligned} \tag{4.4}$$

(6 components, 6 basis 2-forms).

A 1-form is a machine to produce a number out of a vector (bongs of a bell as the vector pierces successive surfaces). A 2-form is a machine to produce a number out of an oriented surface (surface with a sense of circulation indicated on it: Figure 4.1, lower right). The meaning is as clear here as it is in elementary magnetism:

A 2-form as a machine to produce a number out of an oriented surface

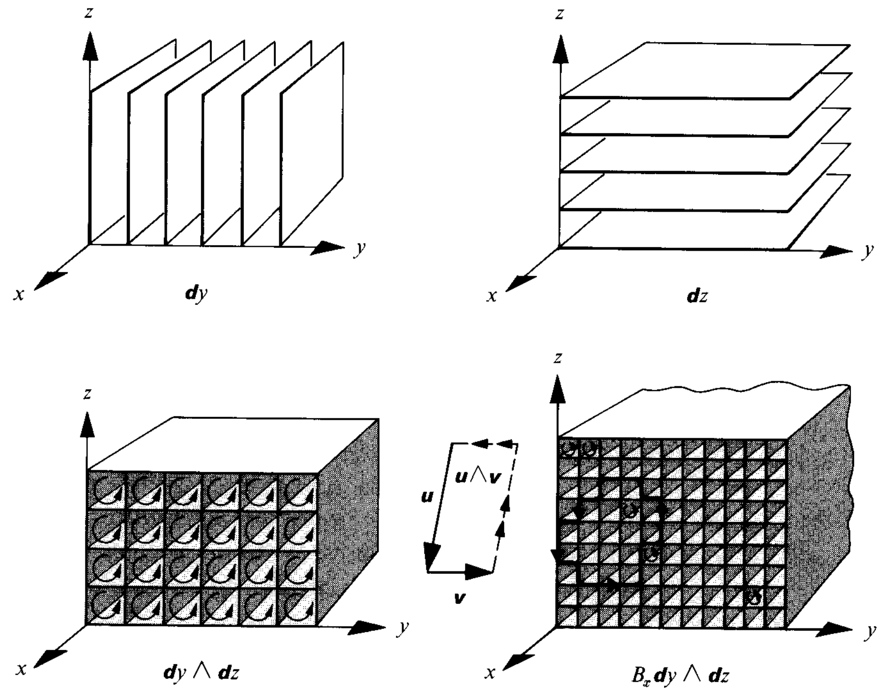


Figure 4.1.

Construction of the 2-form for the electromagnetic field $\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z$ out of the 1-forms $\boldsymbol{d}y$ and $\boldsymbol{d}z$ by "wedge multiplication" (formation of honeycomb-like structure with sense of circulation indicated by arrows). A 2-form is a "machine to construct a number out of an oriented surface" (illustrated by sample surface enclosed by arrows at lower right; number of tubes intersected by this surface is

$$\int_{\text{(this surface)}} \boldsymbol{F} = 18;$$

Faraday's concept of "magnetic flux"). This idea of 2-form machinery can be connected to the "tensor-as-machine" idea of Chapter 3 as follows. The shape of the oriented surface over which one integrates $\boldsymbol{F}$ does not matter, for small surfaces. All that affects $\int \boldsymbol{F}$ is the area of the surface, and its orientation. Choose two vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$, that lie in the surface. They form two legs of a parallelogram, whose orientation ($\boldsymbol{u}$ followed by $\boldsymbol{v}$) and area are embodied in the exterior product $\boldsymbol{u} \wedge \boldsymbol{v}$. Adjust the lengths of $\boldsymbol{u}$ and $\boldsymbol{v}$ so their parallelogram, $\boldsymbol{u} \wedge \boldsymbol{v}$, has the same area as the surface of integration. Then

$$\underbrace{\int_{\text{surface}} \boldsymbol{F} = \int_{\boldsymbol{u}\wedge\boldsymbol{v}} \boldsymbol{F}}_{\substack{\text{machinery idea} \\ \text{of this chapter}}} = \underbrace{\boldsymbol{F}(\boldsymbol{u}, \boldsymbol{v})}_{\substack{\text{machinery idea} \\ \text{of Chapter 3}}}.$$

Exercise: derive this result, for an infinitesimal surface $\boldsymbol{u} \wedge \boldsymbol{v}$ and for general $\boldsymbol{F}$, using the formalism of Box 4.1.

the number of Faraday tubes cut by that surface. The electromagnetic 2-form $\boldsymbol{F}$ or ***Faraday*** described by such a "tubular structure" (suitably abstracted; Box 4.2) has a reality and a location in space that is independent of all coordinate systems and all artificial distinctions between "electric" and "magnetic" fields. Moreover, those tubes provide the most direct geometric representation that anyone has ever been able to give for the machinery by which the electromagnetic field acts on a charged particle. Take a particle of charge e and 4-velocity

$$\boldsymbol{u} = \frac{dx^\alpha}{d\tau}\boldsymbol{e}_\alpha. \tag{4.5}$$

Let this particle go through a region where the electromagnetic field is described by the 2-form

$$\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z \tag{4.6}$$

Lorentz force as contraction of electromagnetic 2-form with particle's 4-velocity

of Figure 4.1. Then the force exerted on the particle (regarded as a 1-form) is the contraction of this 2-form with the 4-velocity (and the charge);

$$\dot{\boldsymbol{p}} = \boldsymbol{d}\boldsymbol{p}/d\tau = e\boldsymbol{F}(\boldsymbol{u}) \equiv e\langle \boldsymbol{F}, \boldsymbol{u}\rangle, \tag{4.7}$$

as one sees by direct evaluation, letting the two factors in the 2-form act in turn on the tangent vector $\boldsymbol{u}$:

$$\begin{aligned}\dot{\boldsymbol{p}} &= eB_x\langle \boldsymbol{d}y \wedge \boldsymbol{d}z, \boldsymbol{u}\rangle \\ &= eB_x\{\boldsymbol{d}y\langle \boldsymbol{d}z, \boldsymbol{u}\rangle - \boldsymbol{d}z\langle \boldsymbol{d}y, \boldsymbol{u}\rangle\} \\ &= eB_x\{\boldsymbol{d}y\langle \boldsymbol{d}z, u^z\boldsymbol{e}_z\rangle - \boldsymbol{d}z\langle \boldsymbol{d}y, u^y\boldsymbol{e}_y\rangle\}\end{aligned}$$

or

$$\dot{p}_\alpha\, \boldsymbol{d}x^\alpha = eB_x u^z\, \boldsymbol{d}y - eB_x u^y\, \boldsymbol{d}z. \tag{4.8}$$

Comparing coefficients of the separate basis 1-forms on the two sides of this equation, one sees reproduced all the detail of the Lorentz force exerted by the magnetic field B_x:

$$\begin{aligned}\dot{p}_y &= \frac{dp_y}{d\tau} = eB_x\frac{dz}{d\tau}, \\ \dot{p}_z &= \frac{dp_z}{d\tau} = -eB_x\frac{dy}{d\tau}.\end{aligned} \tag{4.9}$$

By simple extension of this line of reasoning to the general electromagnetic field, one concludes that *the time-rate of change of momentum (1-form) is equal to the charge multiplied by the contraction of the* ***Faraday*** *with the 4-velocity*. Figure 4.2 illustrates pictorially how the 2-form, $\boldsymbol{F}$, serves as a machine to produce the 1-form, $\dot{\boldsymbol{p}}$, out of the tangent vector, $e\boldsymbol{u}$.

(continued on page 105)

Box 4.2 ABSTRACTING A 2-FORM FROM THE CONCEPT OF "HONEYCOMB-LIKE STRUCTURE," IN 3-SPACE AND IN SPACETIME

Open up a cardboard carton containing a dozen bottles, and observe the honeycomb structure of intersecting north-south and east-west cardboard separators between the bottles. That honeycomb structure of "tubes" ("channels for bottles") is a fairly apt illustration of a 2-form in the context of everyday 3-space. It yields a number (number of tubes cut) for each choice of smooth element of 2-surface slicing through the three-dimensional structure. However, the intersecting cardboard separators are rather too specific. All that a true 2-form can ever give is the number of tubes sliced through, not the "shape" of the tubes. Slew the carton around on the floor by 45°. Then half the separators run NW-SE and the other half run NE-SW, but through a given bit of 2-surface fixed in 3-space the count of tubes is unchanged. Therefore, one should be careful to make the concept of tubes in the mind's eye abstract enough that one envisages direction of tubes (vertical in the example) and density of tubes, but not any specific location or orientation for the tube walls. Thus all the following representations give one and the same 2-form, $\boldsymbol{\sigma}$:

$$\boldsymbol{\sigma} = B\,\boldsymbol{d}x \wedge \boldsymbol{d}y;$$

$$\boldsymbol{\sigma} = B(2\,\boldsymbol{d}x) \wedge \left(\frac{1}{2}\,\boldsymbol{d}y\right)$$

(NS cardboards spaced twice as close as before; EW cardboards spaced twice as wide as before);

$$\boldsymbol{\sigma} = B\,\boldsymbol{d}\left(\frac{x-y}{\sqrt{2}}\right) \wedge \boldsymbol{d}\left(\frac{x+y}{\sqrt{2}}\right)$$

(cardboards rotated through 45°);

$$\boldsymbol{\sigma} = B\frac{\alpha\,\boldsymbol{d}x + \beta\,\boldsymbol{d}y}{(\alpha\delta - \beta\gamma)^{1/2}} \wedge \frac{\gamma\,\boldsymbol{d}x + \delta\,\boldsymbol{d}y}{(\alpha\delta - \beta\gamma)^{1/2}}$$

(both orientation and spacing of "cardboards" changing from point to point, with all four functions, α, β, γ, and δ, depending on position).

What has physical reality, and constitutes the real geometric object, is not any one of the 1-forms just encountered individually, but only the 2-form $\boldsymbol{\sigma}$ itself. This circumstance helps to explain why in the physical literature one sometimes refers to "tubes of force" and sometimes to "lines of force." The two terms for the same structure have this in common, that each yields a number when sliced by a bit of surface. The line-of-force picture has the advantage of not imposing on the mind any specific structure of "sheets of cardboard"; that is, any specific decomposition of the 2-form into the product of 1-forms. However, that very feature is also a disadvantage, for in a calculation one often finds it useful to have a well-defined representation of the 2-form as the wedge product of 1-forms. Moreover, the tube picture, abstract though it must be if it is to be truthful, also has this advantage, that the orientation of the elementary tubes (sense of circulation as indicated by arrows in Figures 4.1 and 4.5, for example) lends itself to ready visualization. Let the "walls" of the tubes therefore remain in all pictures drawn in this book as a reminder that 2-forms can be built out of 1-forms; but let it be understood here and hereafter how manyfold are the options for the individual 1-forms!

Turn now from three dimensions to four, and find that the concept of "honeycomb-like structure" must be made still more abstract. In three dimensions the arbitrariness of the decomposition of the 2-form into 1-forms showed in the slant and packing of the "cardboards," but had no effect on the verticality of the "channels for the bottles" ("direction of Faraday lines of force or tubes of force"); not so in four dimensions, or at least not in the generic case in four dimensions.

In special cases, the story is almost as simple in four dimensions as in three. An example of a special case is once again the 2-form $\boldsymbol{\sigma} = B\,\boldsymbol{d}x \wedge \boldsymbol{d}y$, with all the options for decomposition into 1-forms that have already been mentioned, but with every option giving the same "direction" for the tubes. If the word "direction" now rises in status from "tube walls unpierced by motion in the direction of increasing z" to "tube walls unpierced either by motion in the direction of increasing z, or by motion in the direction of increasing t, or by any linear combination of such motions," that is a natural enough consequence of adding the new dimension. Moreover, the same simplicity prevails for an electromagnetic plane wave. For example, let the wave be advancing in the z-direction, and let the electric polarization point in the x-direction; then for a monochromatic wave, one has

$$E_x = B_y = E_0 \cos \omega(z - t) = -F_{01} = F_{31},$$

and all components distinct from these equal zero. ***Faraday*** is

$$\begin{aligned} \boldsymbol{F} &= F_{01}\,\boldsymbol{d}t \wedge \boldsymbol{d}x + F_{31}\,\boldsymbol{d}z \wedge \boldsymbol{d}x \\ &= E_0 \cos \omega(z - t)\,\boldsymbol{d}(z - t) \wedge \boldsymbol{d}x, \end{aligned}$$

which is again representable as a single wedge product of two 1-forms.

Not so in general! The general 2-form in four dimensions consists of six distinct wedge products,

$$\boldsymbol{F} = F_{01}\,\boldsymbol{d}t \wedge \boldsymbol{d}x + F_{02}\,\boldsymbol{d}t \wedge \boldsymbol{d}y + \cdots + F_{23}\,\boldsymbol{d}y \wedge \boldsymbol{d}z.$$

It is too much to hope that this expression will reduce in the generic case to a single wedge product of two 1-forms (*"simple"* 2-form). It is not even true that it will. It is only remarkable that it can be reduced from six exterior products to two (details in exercise 4.1); thus,

$$\boldsymbol{F} = \boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1 + \boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2.$$

Each product $\boldsymbol{n}^i \wedge \boldsymbol{\xi}^i$ individually can be visualized as a honeycomb-like structure like those depicted in Figures 4.1, 4.2, 4.4, and 4.5. Each such structure individually can be pictured as built out of intersecting sheets (1-forms), but with such details as the tilt and packing of these 1-forms abstracted away. Each such structure individually gives a number when sliced by an element of surface. What counts for the 2-form $\boldsymbol{F}$, however, is neither the number of tubes of $\boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1$ cut by the surface, nor the number of tubes of $\boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2$ cut by the surface, but only the sum of the two. This sum is what is referred to in the text as the "number of tubes of $\boldsymbol{F}$" cut by the surface. The contribution of either wedge product individually is not well-defined, for a simple reason: the decomposition of a six-wedge-product object into two wedge products, miraculous though it seems, is actually far from unique (details in exercise 4.2).

In keeping with the need to have two products of 1-forms to represent the general 2-form note that the vanishing of $\boldsymbol{dF}$ ("no magnetic charges") does not automatically imply that $\boldsymbol{d}(\boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1)$ or $\boldsymbol{d}(\boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2)$ separately vanish. Note also that any spacelike slice through the general 2-form $\boldsymbol{F}$ (reduction from four dimensions to three) can always be represented in terms of a honeycomb-like structure ("simple" 2-form in three dimensions; Faraday's picture of magnetic tubes of force).

Despite the abstraction that has gone on in seeing in all generality what a 2-form is, there is no bar to continuing to use the term "honeycomb-like structure" in a broadened sense to describe this object; and that is the practice here and hereafter.

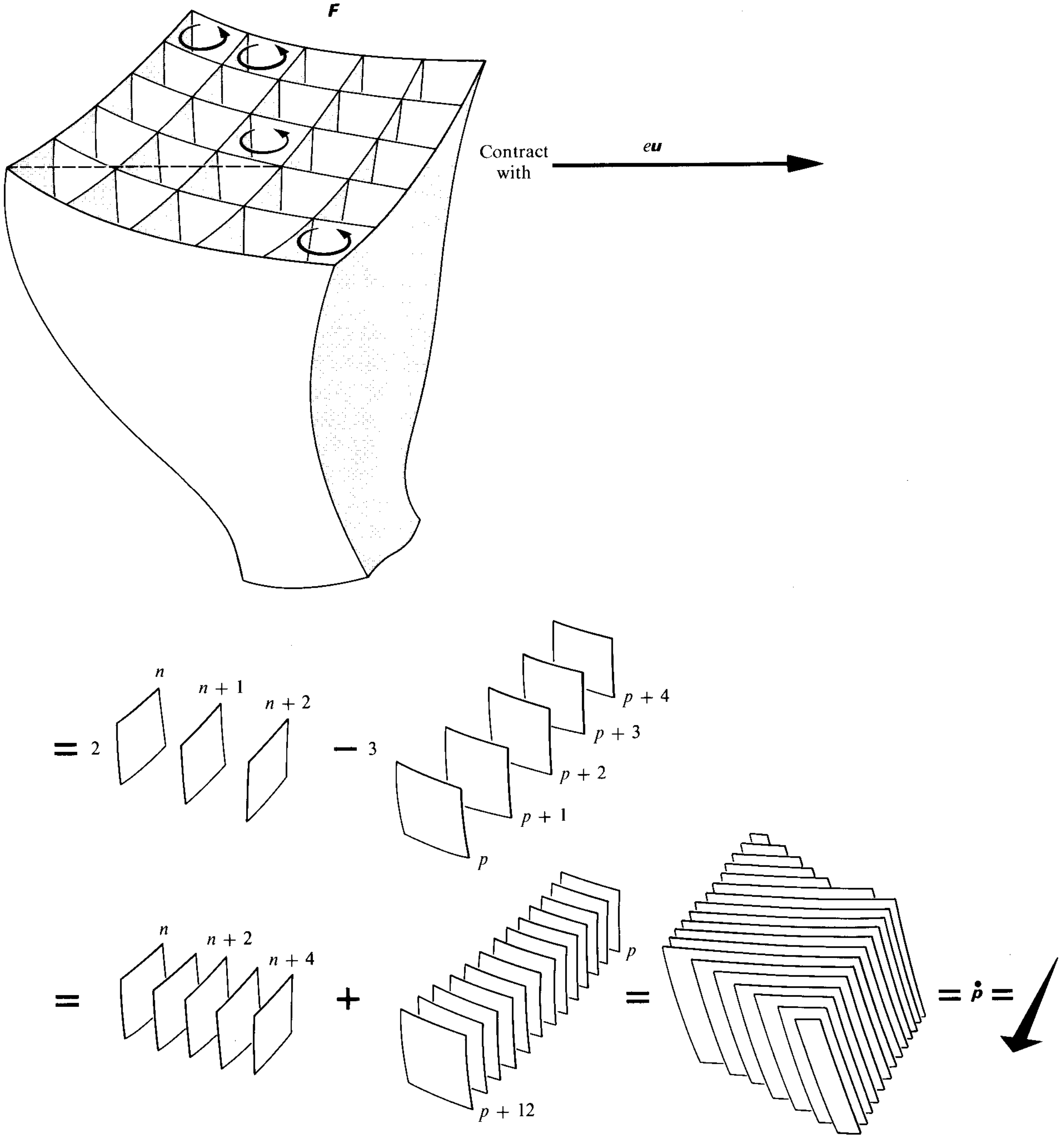


Figure 4.2.
The ***Faraday*** or 2-form $\boldsymbol{F}$ of the electromagnetic field is a machine to produce a 1-form (the time-rate of change of momentum $\dot{\boldsymbol{p}}$ of a charged particle) out of a tangent vector (product of charge e of the particle and its 4-velocity $\boldsymbol{u}$). In spacetime the general 2-form is the "superposition" (see Box 4.2) of two structures like that illustrated at the top of this diagram, the tubes of the first being tilted and packed as indicated, the tubes of the second being tilted in another direction and having a different packing density.

§4.3. FORMS ILLUMINATE ELECTROMAGNETISM, AND ELECTROMAGNETISM ILLUMINATES FORMS

All electromagnetism allows itself to be summarized in the language of 2-forms, honeycomb-like "structures" (again in the abstract sense of "structure" of Box 4.2) of tubes filling all spacetime, as well when spacetime is curved as when it is flat. In brief, there are two such structures, one ***Faraday*** $= \boldsymbol{F}$, the other ***Maxwell*** $= {}^*\boldsymbol{F}$, each dual ("perpendicular," the only place where metric need enter the discussion) to the other, each satisfying an elementary equation:

$$\boldsymbol{dF} = 0 \tag{4.10}$$

("no tubes of ***Faraday*** ever end") and

$$\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J} \tag{4.11}$$

("the number of tubes of ***Maxwell*** that end in an elementary volume is equal to the amount of electric charge in that volume"). To see in more detail how this machinery shows up in action, look in turn at: (1) the definition of a 2-form; (2) the appearance of a given electromagnetic field as ***Faraday*** and as ***Maxwell***; (3) the ***Maxwell*** structure for a point-charge at rest; (4) the same for a point-charge in motion; (5) the nature of the field of a charge that moves uniformly except during a brief instant of acceleration; (6) the ***Faraday*** structure for the field of an oscillating dipole; (7) the concept of exterior derivative; (8) Maxwell's equations in the language of forms; and (9) the solution of Maxwell's equations in flat spacetime, using a 1-form $\boldsymbol{A}$ from which the Liénard-Wiechert 2-form $\boldsymbol{F}$ can be calculated via $\boldsymbol{F} = \boldsymbol{dA}$.

Preview of key points in electromagnetism

A 2-form, as illustrated in Figure 4.1, is a machine to construct a number ("net number of tubes cut") out of any "oriented 2-surface" (2-surface with "sense of circulation" marked on it):

A 2-form as machine for number of tubes cut

$$\begin{pmatrix}\text{number}\\ \text{of tubes}\\ \text{cut}\end{pmatrix} = \int_{\text{surface}} \boldsymbol{F}. \tag{4.12}$$

For example, let the 2-form be the one illustrated in Figure 4.1

Number of tubes cut calculated in one example

$$\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z,$$

and let the surface of integration be the portion of the surface of the 2-sphere $x^2 + y^2 + z^2 = a^2$, $t =$ constant, bounded between $\theta = 70°$ and $\theta = 110°$ and between $\varphi = 0°$ and $\varphi = 90°$ ("Atlantic region of the tropics"). Write

$$\begin{aligned} y &= a \sin\theta \sin\varphi, \\ z &= a\cos\theta, \\ \boldsymbol{d}y &= a\,(\cos\theta \sin\varphi\, \boldsymbol{d}\theta + \sin\theta\cos\varphi\, \boldsymbol{d}\varphi), \\ \boldsymbol{d}z &= -a\sin\theta\, \boldsymbol{d}\theta, \\ \boldsymbol{d}y \wedge \boldsymbol{d}z &= a^2 \sin^2\theta \cos\varphi\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\varphi. \end{aligned} \tag{4.13}$$

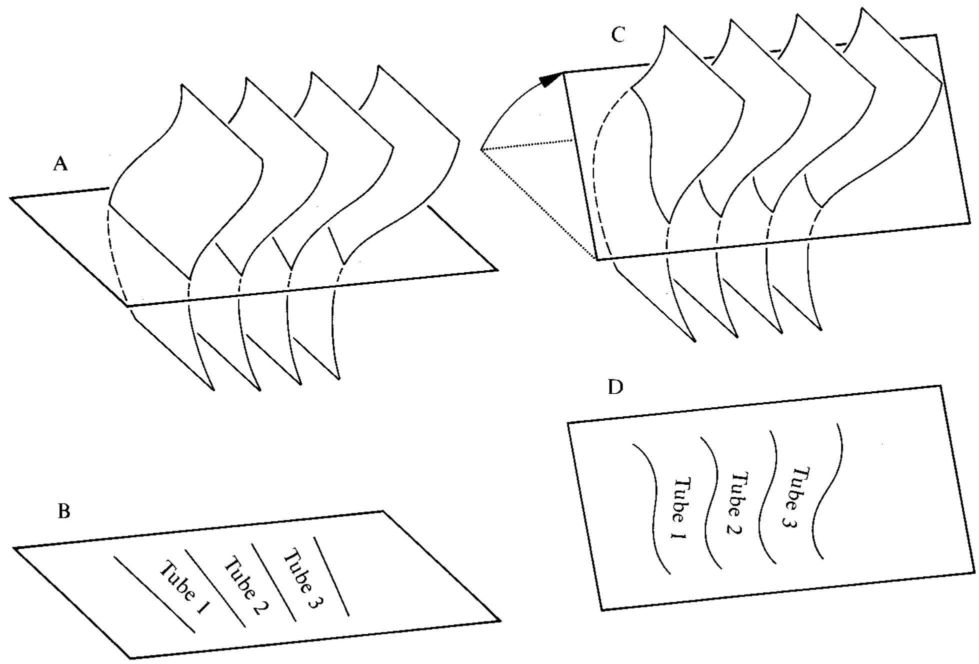


Figure 4.3.
Spacelike slices through ***Faraday***, the electromagnetic 2-form, a geometric object, a honeycomb of tubes that pervades all spacetime ("honeycomb" in the abstract sense spelled out more precisely in Box 4.2). The surfaces in the drawing do not look like a 2-form (honeycomb), because the second family of surfaces making up the honeycomb extends in the spatial direction that is suppressed from the drawing. Diagram A shows one spacelike slice through the 2-form (time increases upwards in the diagram). In diagram B, a projection of the 2-form on this spacelike hypersurface gives the Faraday tubes of magnetic force in this three-dimensional geometry (if the suppressed dimension were restored, the tubes would be tubes, not channels between lines). Diagram C shows another spacelike slice (hypersurface of simultaneity for an observer in a different Lorentz frame). Diagram D shows the very different pattern of magnetic tubes in this reference system. The demand that magnetic tubes of force shall not end ($\nabla \cdot \boldsymbol{B} = 0$), repeated over and over for every spacelike slice through ***Faraday***, gives everywhere the result $\partial \boldsymbol{B}/\partial t = -\nabla \times \boldsymbol{E}$. Thus (magnetostatics) + (covariance) $\longrightarrow$ (magnetodynamics). Similarly—see Chapters 17 and 21—(geometrostatics) + (covariance) $\longrightarrow$ (geometrodynamics).

The structure $\boldsymbol{d}\theta \wedge \boldsymbol{d}\theta$ looks like a "collapsed egg-crate" (Figure 1.4, upper right) and has zero content, a fact formally evident from the vanishing of $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = -\boldsymbol{\beta} \wedge \boldsymbol{\alpha}$ when $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are identical. The result of the integration, assuming constant B_x, is

$$\int_{\text{surface}} \boldsymbol{F} = a^2 B_x \int_{70^\circ}^{110^\circ} \sin^2\theta \, d\theta \int_{0^\circ}^{90^\circ} \cos\varphi \, d\varphi \tag{4.14}$$

It is not so easy to visualize a pure electric field by means of its 2-form $\boldsymbol{F}$ (Figure 4.4, left) as it is to visualize a pure magnetic field by means of its 2-form $\boldsymbol{F}$ (Figures 4.1, 4.2, 4.3). Is there not some way to treat the two fields on more nearly the same footing? Yes, construct the 2-form $^*\boldsymbol{F}$ (Figure 4.4, right) that is *dual* ("perpendicular"; Box 4.3; exercise 3.14) to $\boldsymbol{F}$.

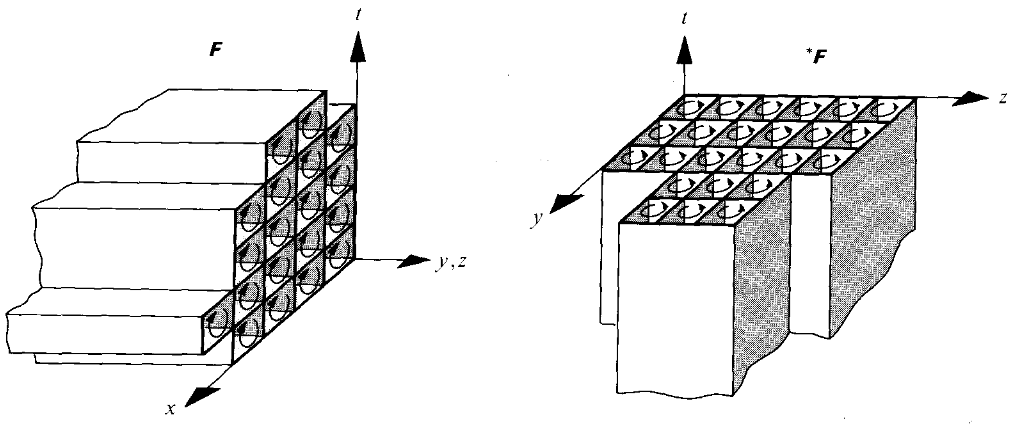


Figure 4.4.
The ***Faraday*** structure

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu} = \frac{1}{2} F_{01}\, \boldsymbol{d}t \wedge \boldsymbol{d}x + \frac{1}{2} F_{10}\, \boldsymbol{d}x \wedge \boldsymbol{d}t = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t$$

associated with an electric field in the x-direction, and the dual ("perpendicular") ***Maxwell*** honeycomb-like 2-form

$$^*\boldsymbol{F} = \frac{1}{2}{}^*F_{\mu\nu}\, \boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu} = {}^*F_{23}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = F^{01}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = F_{10}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = E_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

Represent in geometric form the field of a point-charge of strength e at rest at the origin. Operate in flat space with spherical polar coordinates:

$$\begin{aligned} ds^2 &= -d\tau^2 = g_{\mu\nu}\, dx^{\mu}\, dx^{\nu} \\ &= -dt^2 + dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2. \end{aligned} \tag{4.15}$$

The electric field in the r-direction being $E_r = e/r^2$, it follows that the 2-form $\boldsymbol{F}$ or ***Faraday*** is

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu} = -E_r\, \boldsymbol{d}t \wedge \boldsymbol{d}r = -\frac{e}{r^2}\, \boldsymbol{d}t \wedge \boldsymbol{d}r. \tag{4.16}$$

Its dual, according to the prescription in exercise 3.14, is ***Maxwell***:

Pattern of tubes in dual structure ***Maxwell*** for point-charge at rest

$$\boldsymbol{Maxwell} = {}^*\boldsymbol{F} = e \sin\theta\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\varphi, \tag{4.17}$$

as illustrated in Figure 4.5.

Take a tour in the positive sense around a region of the surface of the sphere illustrated in Figure 4.5. The number of tubes of $^*\boldsymbol{F}$ encompassed in the route will be precisely

$$\begin{pmatrix}\text{number}\\ \text{of tubes}\end{pmatrix} = e \begin{pmatrix}\text{solid}\\ \text{angle}\end{pmatrix}.$$

The whole number of tubes of $^*\boldsymbol{F}$ emergent over the entire sphere will be $4\pi e$, in conformity with Faraday's picture of tubes of force.

Box 4.3 DUALITY OF 2-FORMS IN SPACETIME

Given a general 2-form (containing six exterior or wedge products)

$$\boldsymbol{F} = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + E_y\, \boldsymbol{d}y \wedge \boldsymbol{d}t + \cdots + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y,$$

one gets to its dual ("perpendicular") by the prescription

$$^*\boldsymbol{F} = -B_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots + E_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + E_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y.$$

Duality Rotations

Note that the dual of the dual is the negative of the original 2-form; thus

$$^{**}\boldsymbol{F} = -E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots - B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y = -\boldsymbol{F}.$$

In this sense * has the same property as the imaginary number i: $** = ii = -1$. Thus one can write

$$e^{*\alpha} = \cos\alpha + {}^*\sin\alpha.$$

This operation, applied to $\boldsymbol{F}$, carries attention from the generic 2-form in its simplest representation (see exercise 4.1)

$$\boldsymbol{F} = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z$$

to another "duality rotated electromagnetic field"

$$e^{*\alpha}\boldsymbol{F} = (E_x \cos\alpha - B_x \sin\alpha)\, \boldsymbol{d}x \wedge \boldsymbol{d}t + (B_x \cos\alpha + E_x \sin\alpha)\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

If the original field satisfied Maxwell's empty-space field equations, so does the new field. With suitable choice of the "complexion" α, one can annul one of the two wedge products at any chosen point in spacetime and have for the other

$$(B_x^2 + E_x^2)^{1/2}\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

Field of a point-charge in motion

How can one determine the structure of tubes associated with a charged particle moving at a uniform velocity? First express $^*\boldsymbol{F}$ in rectangular coordinates moving with the particle (barred coordinates in this comoving "rocket" frame of reference; unbarred coordinates will be used later for a laboratory frame of reference). The relevant steps can be listed:

(a)

$$^*\boldsymbol{F} = e \sin\bar{\theta}\, \boldsymbol{d}\bar{\theta} \wedge \boldsymbol{d}\bar{\varphi} = -e(\boldsymbol{d}\cos\bar{\theta}) \wedge \boldsymbol{d}\bar{\varphi};$$

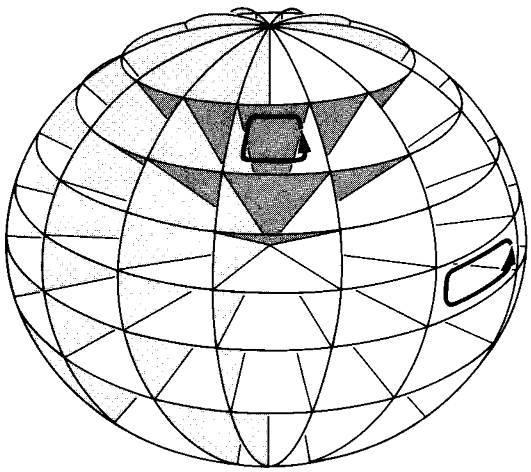


Figure 4.5.
The field of 2-forms ***Maxwell*** $= {}^*\boldsymbol{F} = e \sin\theta\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\phi$ that describes the electromagnetic field of a charge e at rest at the origin. This picture is actually the intersection of ${}^*\boldsymbol{F}$ with a 3-surface of constant time t; i.e., the time direction is suppressed from the picture.

(b)

$$\bar\varphi = \arctan\frac{\bar y}{\bar x}; \qquad \boldsymbol{d}\bar\varphi = \frac{\bar x\, \boldsymbol{d}\bar y - \bar y\, \boldsymbol{d}\bar x}{\bar x^2 + \bar y^2};$$

(c)

$$\cos\bar\theta = \frac{\bar z}{\bar r}; \qquad -\boldsymbol{d}(\cos\bar\theta) = \frac{-\boldsymbol{d}\bar z}{\bar r} + \frac{\bar z}{\bar r^3}(\bar x\, \boldsymbol{d}\bar x + \bar y\, \boldsymbol{d}\bar y + \bar z\, \boldsymbol{d}\bar z);$$

(d) combine to find

$$ {}^*\boldsymbol{F} = (e/\bar r^3)(\bar x\, \boldsymbol{d}\bar y \wedge \boldsymbol{d}\bar z + \bar y\, \boldsymbol{d}\bar z \wedge \boldsymbol{d}\bar x + \bar z\, \boldsymbol{d}\bar x \wedge \boldsymbol{d}\bar y) \qquad (4.18)$$

(electromagnetic field of point charge in a comoving Cartesian system; spherically symmetric). Now transform to laboratory coordinates:

$$\text{velocity parameter } \alpha$$

$$\text{velocity } \beta = \tanh\alpha$$

$$\frac{1}{\sqrt{1-\beta^2}} = \cosh\alpha, \qquad \frac{\beta}{\sqrt{1-\beta^2}} = \sinh\alpha$$

(a)
$$\begin{cases} \bar t = t\cosh\alpha - x\sinh\alpha, \\ \bar x = -t\sinh\alpha + x\cosh\alpha, \\ \bar y = y \qquad \bar z = z; \end{cases}$$

(b)
$$\bar r = [(x\cosh\alpha - t\sinh\alpha)^2 + y^2 + z^2]^{1/2};$$

(c)
$$ {}^*\boldsymbol{F} = (e/\bar r^3)[(x\cosh\alpha - t\sinh\alpha)\, \boldsymbol{d}y \wedge \boldsymbol{d}z + y\, \boldsymbol{d}z \wedge (\cosh\alpha\, \boldsymbol{d}x - \sinh\alpha\, \boldsymbol{d}t) + z(\cosh\alpha\, \boldsymbol{d}x - \sinh\alpha\, \boldsymbol{d}t) \wedge \boldsymbol{d}y]; \qquad (4.19)$$

(d) compare with the general dual 2-form,

$$\begin{aligned} {}^*\boldsymbol{F} = E_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z + E_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + E_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y \\ + B_x\, \boldsymbol{d}t \wedge \boldsymbol{d}x + B_y\, \boldsymbol{d}t \wedge \boldsymbol{d}y + B_z\, \boldsymbol{d}t \wedge \boldsymbol{d}z; \end{aligned}$$

and get the desired individual field components

(e)
$$\begin{cases} E_x = (e/\bar{r}^3)(x \cosh \alpha - t \sinh \alpha), & B_x = 0, \\ E_y = (e/\bar{r}^3) y \cosh \alpha, & B_y = -(e/\bar{r}^3) z \sinh \alpha, \\ E_z = (e/\bar{r}^3) z \cosh \alpha, & B_z = (e/\bar{r}^3) y \sinh \alpha. \end{cases} \tag{4.20}$$

One can verify that the invariants

$$\boldsymbol{B}^2 - \boldsymbol{E}^2 = \frac{1}{2} F_{\alpha\beta} F^{\alpha\beta}, \tag{4.21}$$

$$\boldsymbol{E} \cdot \boldsymbol{B} = \frac{1}{4} F_{\alpha\beta} {}^*F^{\alpha\beta} \tag{4.22}$$

have the same value in the laboratory frame as in the rocket frame, as required. Note that the honeycomb structure of the differential form is not changed when one goes from the rocket frame to the laboratory frame. What changes is only the mathematical formula that describes it.

§4.4. RADIATION FIELDS

How an acceleration causes radiation

The ***Maxwell*** structure of tubes associated with a charge in uniform motion is more remarkable than it may seem at first sight, and not only because of the Lorentz contraction of the tubes in the direction of motion. The tubes arbitrarily far away move on in military step with the charge on which they center, despite the fact that there is no time for information "emitted" from the charge "right now" to get to the faraway tube "right now." The structure of the faraway tubes "right now" must therefore derive from the charge at an earlier moment on its uniform-motion, straight-line trajectory. This circumstance shows up nowhere more clearly than in what happens to the field in consequence of a sudden change, in a short time $\Delta\tau$, from one uniform velocity to another uniform velocity (Figure 4.6). The tubes have the standard patterns for the two states of motion, one pattern within a sphere of radius r, the other outside that sphere, where r is equal to the lapse of time ("cm of light-travel time") since the acceleration took place. The necessity for the two patterns to fit together in the intervening zone, of thickness $\Delta r = \Delta\tau$, forces the field there to be multiplied up by a "stretching factor," proportional to r. This factor is responsible for the well-known fact that radiative forces fall off inversely only as the first power of the distance (Figure 4.6).

When the charge continuously changes its state of motion, the structure of the electromagnetic field, though based on the same simple principles as those illustrated in Figure 4.6, nevertheless looks more complex. The following is the ***Faraday*** 2-form

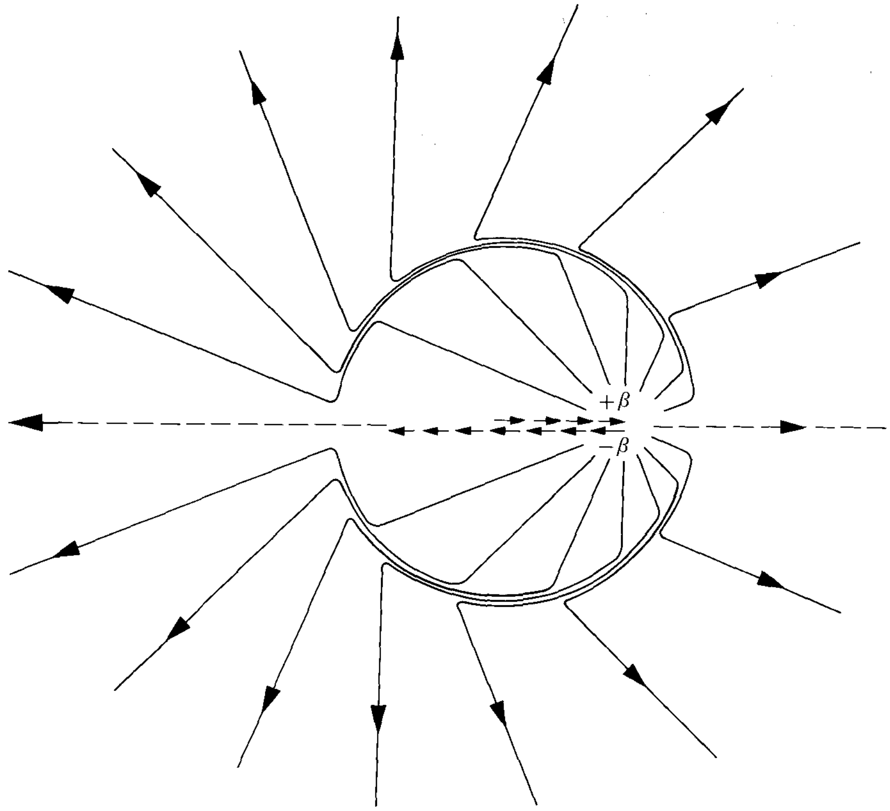


Figure 4.6.
Mechanism of radiation. J. J. Thomson's way to understand why the strength of an electromagnetic wave falls only as the inverse first power of distance r and why the amplitude of the wave varies (for low velocities) as $\sin\theta$ (maximum in the plane perpendicular to the line of acceleration). The charge was moving to the left at uniform velocity. Far away from it, the lines of force continue to move as if this uniform velocity were going to continue forever (Coulomb field of point-charge in slow motion). However, closer up the field is that of a point-change moving to the right with uniform velocity ($1/r^2$ dependence of strength upon distance). The change from the one field pattern to another is confined to a shell of thickness $\Delta\tau$ located at a distance r from the point of acceleration (amplification of field by "stretching factor" $r\sin\theta\,\Delta\beta/\Delta\tau$; see text). We thank C. Teitelboim for the construction of this diagram.

for the field of an electric dipole of magnitude p_1 oscillating up and down parallel to the z-axis:

Field of an oscillating dipole

$$\begin{aligned}\boldsymbol{F} = E_x\,\boldsymbol{d}x \wedge \boldsymbol{d}t + \cdots + B_x\,\boldsymbol{d}y \wedge \boldsymbol{d}z + \cdots = \text{real part of } \{p_1 e^{i\omega r - i\omega t} \\ \underbrace{[2\cos\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2}\right)}_{\text{gives } E_r} \boldsymbol{d}r \wedge \boldsymbol{d}t + \underbrace{\sin\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } E_\theta} r\,\boldsymbol{d}\theta \wedge \boldsymbol{d}t \\ \underbrace{+ \sin\theta\left(\frac{-i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } B_\phi} \boldsymbol{d}r \wedge r\,\boldsymbol{d}\theta]\}\end{aligned} \tag{4.23}$$

and the dual 2-form ***Maxwell*** = ****F*** is

$$
\begin{aligned}
{}^{*}\boldsymbol{F} = -B_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots + E_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z + \cdots = \text{real part of } \{p_1 e^{i\omega r - i\omega t} \\
\underbrace{[\sin\theta\left(\frac{-i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } B_\phi} \boldsymbol{d}t \wedge r\sin\theta\, \boldsymbol{d}\phi \\
\underbrace{+\, 2\cos\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2}\right)}_{\text{gives } E_r} r\, \boldsymbol{d}\theta \wedge r\sin\theta\, \boldsymbol{d}\phi \\
\underbrace{+\, \sin\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } E_\theta} r\sin\theta\, \boldsymbol{d}\phi \wedge \boldsymbol{d}r]\}.
\end{aligned}
\tag{4.24}
$$

§4.5. MAXWELL'S EQUATIONS

The general 2-form $\boldsymbol{F}$ is written as a superposition of wedge products with a factor $\frac{1}{2}$,

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu, \tag{4.25}$$

because the typical term appears twice, once as $F_{xy}\, \boldsymbol{d}x \wedge \boldsymbol{d}y$ and the second time as $F_{yx}\, \boldsymbol{d}y \wedge \boldsymbol{d}x$, with $F_{yx} = -F_{xy}$ and $\boldsymbol{d}y \wedge \boldsymbol{d}x = -\boldsymbol{d}x \wedge \boldsymbol{d}y$.

If differentiation ("taking the gradient"; the operator $\boldsymbol{d}$) produced out of a scalar a 1-form, it is also true that differentiation (again the operator $\boldsymbol{d}$, but now generally known under Cartan's name of "exterior differentiation") produces a 2-form out of the general 1-form; and applied to a 2-form produces a 3-form; and applied to a 3-form produces a 4-form, the form of the highest order that spacetime will accommodate. Write the general f-form as

$$\boldsymbol{\phi} = \frac{1}{f!}\, \phi_{\alpha_1\alpha_2\ldots\alpha_f}\, \boldsymbol{d}x^{\alpha_1} \wedge \boldsymbol{d}x^{\alpha_2} \wedge \cdots \wedge \boldsymbol{d}x^{\alpha_f} \tag{4.26}$$

where the coefficient $\phi_{\alpha_1\alpha_2\ldots\alpha_f}$, like the wedge product that follows it, is antisymmetric under interchange of any two indices. Then the exterior derivative of $\boldsymbol{\phi}$ is

Taking exterior derivative

$$\boldsymbol{d\phi} \equiv \frac{1}{f!}\frac{\partial\phi_{\alpha_1\alpha_2\ldots\alpha_f}}{\partial x^{\alpha_0}}\, \boldsymbol{d}x^{\alpha_0} \wedge \boldsymbol{d}x^{\alpha_1} \wedge \boldsymbol{d}x^{\alpha_2} \wedge \cdots \wedge \boldsymbol{d}x^{\alpha_f}. \tag{4.27}$$

Take the exterior derivative of ***Faraday*** according to this rule and find that it vanishes, not only for the special case of the dipole oscillator, but also for a general electromagnetic field. Thus, in the coordinates appropriate for a local Lorentz frame, one has

$$\begin{aligned}\boldsymbol{dF} &= \boldsymbol{d}(E_x\, \boldsymbol{dx} \wedge \boldsymbol{dt} + \cdots + B_x\, \boldsymbol{dy} \wedge \boldsymbol{dz} + \cdots)\\ &= \left(\frac{\partial E_x}{\partial t}\boldsymbol{dt} + \frac{\partial E_x}{\partial x}\boldsymbol{dx} + \frac{\partial E_x}{\partial y}\boldsymbol{dy} + \frac{\partial E_x}{\partial z}\boldsymbol{dz}\right) \wedge \boldsymbol{dx} \wedge \boldsymbol{dt}\\ &\quad + \cdots \text{(5 more such sets of 4 terms each)} \ldots . \end{aligned} \tag{4.28}$$

Note that such a term as $\boldsymbol{dy} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz}$ is automatically zero ("collapse of egg-crate cell when stamped on"). Collect the terms that do not vanish and find

$$\begin{aligned}\boldsymbol{dF} = &\left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}\right)\boldsymbol{dx} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz}\\ &+ \left(\frac{\partial B_x}{\partial t} + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\boldsymbol{dt} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz}\\ &+ \left(\frac{\partial B_y}{\partial t} + \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\boldsymbol{dt} \wedge \boldsymbol{dz} \wedge \boldsymbol{dx}\\ &+ \left(\frac{\partial B_z}{\partial t} + \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\boldsymbol{dt} \wedge \boldsymbol{dx} \wedge \boldsymbol{dy}. \end{aligned} \tag{4.29}$$

Each term in this expression is familiar from Maxwell's equations

$$\operatorname{div} \boldsymbol{B} = \boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$$

and

$$\operatorname{curl} \boldsymbol{E} = \boldsymbol{\nabla} \times \boldsymbol{E} = -\dot{\boldsymbol{B}}.$$

Each vanishes, and with their vanishing ***Faraday*** itself is seen to have zero exterior derivative:

$$\boldsymbol{dF} = 0. \tag{4.30}$$

In other words, "***Faraday*** is a closed 2-form"; "the tubes of $\boldsymbol{F}$ nowhere come to an end."

Faraday structure: tubes nowhere end

A similar calculation gives for the exterior derivative of the dual 2-form ***Maxwell*** the result

$$\begin{aligned}\boldsymbol{d}{}^*\boldsymbol{F} &= \boldsymbol{d}(-B_x\, \boldsymbol{dx} \wedge \boldsymbol{dt} - \cdots + E_x\, \boldsymbol{dy} \wedge \boldsymbol{dz} + \cdots)\\ &= \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right)\boldsymbol{dx} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz}\\ &\quad + \left(\frac{\partial E_x}{\partial t} - \frac{\partial B_z}{\partial y} + \frac{\partial B_y}{\partial z}\right)\boldsymbol{dt} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz}\\ &\quad + \cdots\\ &= 4\pi(\rho\, \boldsymbol{dx} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz}\\ &\qquad - J_x\, \boldsymbol{dt} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz}\\ &\qquad - J_y\, \boldsymbol{dt} \wedge \boldsymbol{dz} \wedge \boldsymbol{dx}\\ &\qquad - J_z\, \boldsymbol{dt} \wedge \boldsymbol{dx} \wedge \boldsymbol{dy}) = 4\pi\, {}^*\boldsymbol{J};\end{aligned}$$

Maxwell structure: density of tube endings given by charge-current 3-form

$$\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J}. \tag{4.31}$$

In empty space this exterior derivative, too, vanishes; there ***Maxwell*** is a closed 2-form; the tubes of $^*\boldsymbol{F}$, like the tubes of $\boldsymbol{F}$, nowhere come to an end.

In a region where charge is present, the situation changes. Tubes of ***Maxwell*** take their origin in such a region. The density of endings is described by the 3-form $^*\boldsymbol{J}$ = ***charge***, a "collection of eggcrate cells" collected along bundles of world lines.

The two equations

$$\boldsymbol{dF} = 0$$

and

$$\boldsymbol{d}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J}$$

summarize the entire content of Maxwell's equations in geometric language. The forms $\boldsymbol{F}$ = ***Faraday***, and $^*\boldsymbol{F}$ = ***Maxwell***, can be described in any coordinates one pleases—or in a language (honeycomb and egg-crate structures) free of any reference whatsoever to coordinates. Remarkably, neither equation makes any reference whatsoever to *metric*. As Hermann Weyl was one of the most emphatic in stressing (see also Chapters 8 and 9), the concepts of form and exterior derivative are metric-free. Metric made an appearance only in one place, in the concept of duality ("perpendicularity") that carried attention from $\boldsymbol{F}$ to the dual structure $^*\boldsymbol{F}$.

Duality: the only place in electromagnetism where metric must enter

§4.6. EXTERIOR DERIVATIVE AND CLOSED FORMS

The words "honeycomb" and "egg crate" may have given some feeling for the geometry that goes with electrodynamics. Now to spell out these concepts more clearly and illustrate in geometric terms, with electrodynamics as subject matter, what it means to speak of "exterior differentiation." Marching around a boundary, yes; but how and why and with what consequences? It is helpful to return to functions and 1-forms, and see them and the 2-forms ***Faraday*** and ***Maxwell*** and the 3-form ***charge*** as part of an ordered progression (see Box 4.4). Two-forms are seen in this box to be of two kinds: (1) a special 2-form, known as a "closed" 2-form, which has the property that as many tubes enter a closed 2-surface as emerge from it (exterior derivative of 2-form zero; no 3-form derivable from it other than the trivial *zero* 3-form!); and (2) a general 2-form, which sends across a closed 2-surface a non-zero net number of tubes, and therefore permits one to define a nontrivial 3-form ("exterior derivative of the 2-form"), which has precisely as many egg-crate cells in any closed 2-surface as the net number of tubes of the 2-form emerging from that same closed 2-surface (generalization of Faraday's concept of tubes of force to the world of spacetime, curved as well as flat).

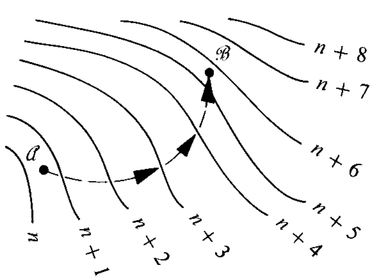
Closed 2-form contrasted with general 2-form

(*continued on page 120*)

Box 4.4 THE PROGRESSION OF FORMS AND EXTERIOR DERIVATIVES

0-Form or Scalar, f

An example in the context of 3-space and Newtonian physics is temperature, $T(x, y, z)$, and in the context of spacetime, a scalar potential, $\phi(t, x, y, z)$.

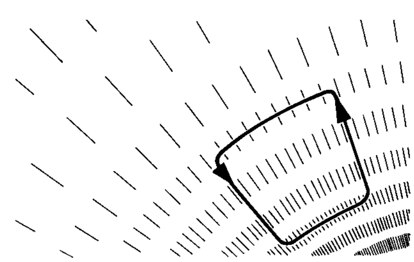


From Scalar to *1*-Form

Take the gradient or "exterior derivative" of a scalar f to obtain a special 1-form, $\boldsymbol{\gamma} = \boldsymbol{d}f$. Comments: (a) Any additive constant included in f is erased in the process of differentiation; the quantity n in the diagram at the left is unknown and irrelevant. (b) The 1-form $\boldsymbol{\gamma}$ is special in the sense that surfaces in one region "mesh" with surfaces in a neighboring region ("closed 1-form"). (c) Line integral $\int_{\mathcal{A}}^{\mathcal{B}} \boldsymbol{d}f$ is independent of path for any class of paths equivalent to one another under continuous deformation. (d) The 1-form is a machine to produce a number ("bongs of bell" as each successive integral surface is crossed) out of a displacement (approximation to concept of a tangent vector).

General *1*-Form $\boldsymbol{\beta} = \beta_\alpha \, \boldsymbol{d}x^\alpha$

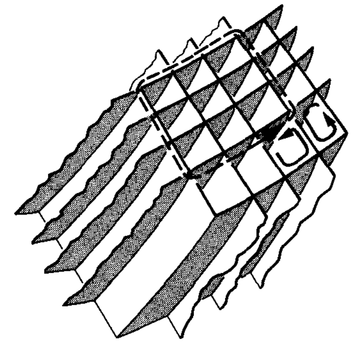
This is a pattern of surfaces, as illustrated in the diagram at the right; i.e., a machine to produce a number ("bongs of bell"; $\langle \boldsymbol{\beta}, \boldsymbol{u} \rangle$) out of a vector. A 1-form has a reality and position in space independent of all choice of coordinates. Surfaces do not ordinarily mesh. Integral $\oint \boldsymbol{\beta}$ around indicated closed loop does not give zero ("more bongs than antibongs").



From *1*-Form to *2*-Form $\boldsymbol{\xi} = \boldsymbol{d}\boldsymbol{\beta} = \dfrac{\partial \beta_\alpha}{\partial x^\mu} \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\alpha$

$\boldsymbol{\xi}$ is a pattern of honeycomb-like cells, with a direction of circulation marked on each, so stationed

Box 4.4 ***(continued)***



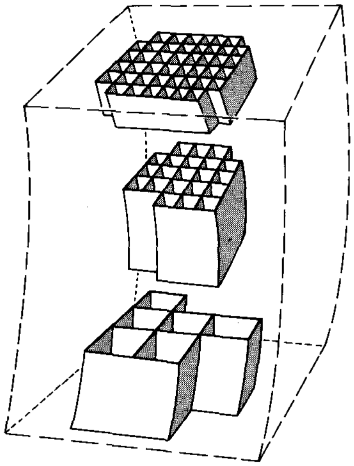
that the number of cells encompassed in the dotted closed path is identical to the net contribution (excess of bongs over antibongs) for the same path in the diagram of $\boldsymbol{\beta}$ above. The "exterior derivative" is *defined* so this shall be so; the generalized Stokes theorem codifies it. The word "exterior" comes from the fact that the path goes around the periphery of the region under analysis. Thus the 2-form is a machine to get a number (number of tubes, $\langle \boldsymbol{\xi}, \boldsymbol{u} \wedge \boldsymbol{v} \rangle$) out of a bit of surface ($\boldsymbol{u} \wedge \boldsymbol{v}$) that has a sense of circulation indicated upon it. The 2-form thus defined is special in this sense: a rubber sheet "supported around its edges" by the dotted curve or any other closed curve is crossed by the same number of tubes when: (a) it bulges up in the middle; (b) it is pushed down in the middle; (c) it experiences any other continuous deformation. The ***Faraday*** or 2-form $\boldsymbol{F}$ of electromagnetism, always expressible as $\boldsymbol{F} = \boldsymbol{dA}$ ($\boldsymbol{A}$ = 4-potential, a 1-form), also has always this special property ("conservation of tubes").

0-Form to *1*-Form to *2*-Form? No!

Go from scalar f to 1-form $\boldsymbol{\gamma} = \boldsymbol{d}f$. The next step to a 2-form $\boldsymbol{\alpha}$ is vacuous. The net contribution of the line integral $\int \boldsymbol{\gamma}$ around the dotted closed path is automatically zero. To reproduce that zero result requires a zero 2-form. Thus $\boldsymbol{\alpha} = \boldsymbol{d\gamma} = \boldsymbol{dd}f$ has to be the zero 2-form. This result is a special instance of the general result $\boldsymbol{dd} = 0$.

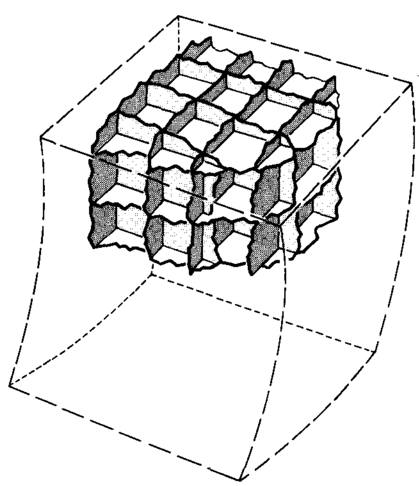
General *2*-Form $\boldsymbol{\sigma} = \frac{1}{2}\sigma_{\alpha\beta}\, \boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta}$, with $\sigma_{\alpha\beta} = -\sigma_{\beta\alpha}$

Again, this is a honeycomb-like structure, and again a machine to get a number (number of tubes, $\langle \boldsymbol{\sigma}, \boldsymbol{u} \wedge \boldsymbol{v} \rangle$) out of a surface ($\boldsymbol{u} \wedge \boldsymbol{v}$) that has a sense of circulation indicated on it. It is general in the sense that the honeycomb structures in one region do not ordinarily mesh with those in a neighboring region. In consequence, a closed 2-surface, such as the box-like surface indicated by dotted lines at the right, is ordinarily crossed by a non-zero net number of tubes. The net number of tubes emerging from such a closed surface is, however, exactly zero when the 2-form is the exterior derivative of a 1-form.



From *2*-Form to *3*-Form $\boldsymbol{\mu} = \boldsymbol{d\sigma} = \dfrac{\partial \sigma_{|\alpha\beta|}}{\partial x^\gamma} \boldsymbol{dx}^\gamma \wedge \boldsymbol{dx}^\alpha \wedge \boldsymbol{dx}^\beta$,
where $\boldsymbol{dx}^\gamma \wedge \boldsymbol{dx}^\alpha \wedge \boldsymbol{dx}^\beta \equiv 3!\, \boldsymbol{dx}^{[\gamma} \otimes \boldsymbol{dx}^\alpha \otimes \boldsymbol{dx}^{\beta]}$

This egg-crate type of structure is a machine to get a number (number of cells $\langle \boldsymbol{\mu}, \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w} \rangle$) from a volume (volume $\boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}$ within which one counts the cells). A more complete diagram would provide each cell and the volume of integration itself with an indicator of orientation (analogous to the arrow of circulation shown for cells of the 2-form). The contribution of a given cell to the count of cells is $+1$ or -1, according as the orientation indicators have same sense or opposite sense. The number of egg-crate cells of $\boldsymbol{\mu} = \boldsymbol{d\sigma}$ in any given volume (such as the volume indicated by the dotted lines) is tailored to give precisely the same number as the net number of tubes of the 2-form $\boldsymbol{\sigma}$ (diagram above) that emerge from that volume (generalized Stokes theorem). For electromagnetism, the exterior derivative of ***Faraday*** or 2-form $\boldsymbol{F}$ gives a null 3-form, but the exterior derivative of ***Maxwell*** or 2-form $^*\boldsymbol{F}$ gives 4π times the 3-form $^*\boldsymbol{J}$ of charge:



$$\begin{aligned} {}^*\boldsymbol{J} = \rho\, \boldsymbol{dx} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz} - J_x\, \boldsymbol{dt} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz} \\ - J_y\, \boldsymbol{dt} \wedge \boldsymbol{dz} \wedge \boldsymbol{dx} - J_z\, \boldsymbol{dt} \wedge \boldsymbol{dx} \wedge \boldsymbol{dy}. \end{aligned}$$

Box 4.4 (*continued*)

From *1*-Form to *2*-Form to *3*-Form? No!

Starting with a 1-form (electromagnetic 4-potential), construct its exterior derivative, the 2-form $\boldsymbol{F} = \boldsymbol{dA}$ (***Faraday***). The tubes in this honeycomb-like structure never end. So the number of tube endings in any elementary volume, and with it the 3-form $\boldsymbol{dF} = \boldsymbol{ddA}$, is automatically zero. This is another example of the general result that $\boldsymbol{dd} = 0$.

From *2*-Form to *3*-Form to *4*-Form? No!

Starting with 2-form $\boldsymbol{{}^*F}$ (***Maxwell***), construct its exterior derivative, the 3-form $4\pi\, \boldsymbol{{}^*J}$. The cells in this egg-crate type of structure extend in a fourth dimension ("hypertube"). The number of these hypertubes that end in any elementary 4-volume, and with it the 4-form

$$\boldsymbol{d}(4\pi\, \boldsymbol{{}^*J}) = \boldsymbol{dd{}^*F},$$

is automatically zero, still another example of the general result that $\boldsymbol{dd} = 0$. This result says that

$$\boldsymbol{d{}^*J} = \left(\frac{\partial \rho}{\partial t} + \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z}\right) \boldsymbol{dt} \wedge \boldsymbol{dx} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz} = 0$$

("law of conservation of charge"). Note:

$$\boldsymbol{dx}^\alpha \wedge \boldsymbol{dx}^\beta \wedge \boldsymbol{dx}^\gamma \wedge \boldsymbol{dx}^\delta \equiv 4!\, \boldsymbol{dx}^{[\alpha} \otimes \boldsymbol{dx}^\beta \otimes \boldsymbol{dx}^\gamma \otimes \boldsymbol{dx}^{\delta]}.$$

This implies $\boldsymbol{dt} \wedge \boldsymbol{dx} \wedge \boldsymbol{dy} \wedge \boldsymbol{dz} = \boldsymbol{\varepsilon}$.

From *3*-Form to *4*-Form $\boldsymbol{\tau} = \boldsymbol{d\nu} = \dfrac{\partial \nu_{|\alpha\beta\gamma|}}{\partial x^\delta} \boldsymbol{dx}^\delta \wedge \boldsymbol{dx}^\alpha \wedge \boldsymbol{dx}^\beta \wedge \boldsymbol{dx}^\gamma$

This four-dimensional "super-egg-crate" type structure is a machine to get a number (number of cells, $\langle \boldsymbol{\tau}, \boldsymbol{n} \wedge \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w} \rangle$) from a 4-volume $\boldsymbol{n} \wedge \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}$.

From 4-Form to 5-Form? No!

Spacetime, being four-dimensional, cannot accommodate five-dimensional egg-crate structures. At least two of the $\boldsymbol{d}x^\mu$'s in

$$\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\delta \wedge \boldsymbol{d}x^\varepsilon$$

must be the same; so, by antisymmetry of "$\wedge$," this "basis 5-form" must vanish.

Results of Exterior Differentiation, Summarized

0-form	f				
1-form	$\boldsymbol{d}f$	$\boldsymbol{A}$			
2-form	$\boldsymbol{dd}f \equiv 0$	$\boldsymbol{F} = \boldsymbol{dA}$	$^*\boldsymbol{F}$		
3-form		$\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$	$4\pi\, ^*\boldsymbol{J} = \boldsymbol{d}^*\boldsymbol{F}$	$\boldsymbol{\nu}$	
4-form			$\boldsymbol{d}(4\pi\, ^*\boldsymbol{J}) = \boldsymbol{dd}^*\boldsymbol{F} \equiv 0$	$\boldsymbol{\tau} = \boldsymbol{d\nu}$	$\boldsymbol{\mu}$
5-form?	No!			$\boldsymbol{d\tau} \equiv 0$	$\boldsymbol{d\mu} \equiv 0$

New Forms from Old by Taking Dual (see exercise 3.14)

Dual of scalar f is 4-form: $^*f = f\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = f\boldsymbol{\varepsilon}$.

Dual of 1-form $\boldsymbol{J}$ is 3-form: $^*\boldsymbol{J} = J^0\,\boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 - J^1\,\boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 \wedge \boldsymbol{d}x^0$
$+ J^2\,\boldsymbol{d}x^3 \wedge \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 - J^3\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2$.

Dual of 2-form $\boldsymbol{F}$ is 2-form: $^*\boldsymbol{F} = F^{|\alpha\beta|}\varepsilon_{\alpha\beta|\mu\nu|}\,\boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu$, where
$F^{\alpha\beta} = \eta^{\alpha\lambda}\eta^{\beta\delta}F_{\lambda\delta}$.

Dual of 3-form $\boldsymbol{K}$ is 1-form: $^*\boldsymbol{K} = K^{012}\,\boldsymbol{d}x^3 - K^{123}\,\boldsymbol{d}x^0 + K^{230}\,\boldsymbol{d}x^1 - K^{301}\,\boldsymbol{d}x^2$,
where $K^{\alpha\beta\gamma} = \eta^{\alpha\mu}\eta^{\beta\nu}\eta^{\gamma\lambda}K_{\mu\nu\lambda}$.

Dual of 4-form $\boldsymbol{L}$ is a scalar: $\boldsymbol{L} = L_{0123}\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3$;
$^*\boldsymbol{L} = L^{0123} = -L_{0123}$.

Note 1: This concept of duality between *one form and another* is to be distinguished from the concept of duality between the *vector basis* $\boldsymbol{e}_\alpha$ and the *1-form basis* $\boldsymbol{\omega}^\alpha$ of a given frame. The two types of duality have nothing whatsoever to do with each other!

Box 4.4 *(continued)*

Note 2: In spacetime, the operation of taking the dual, applied twice, leads back to the original form for forms of odd order, and to the negative thereof for forms of even order. In Euclidean 3-space the operation reproduces the original form, regardless of its order.

Duality Plus Exterior Differentiation

Start with scalar ϕ. Its gradient $\boldsymbol{d}\phi$ is a 1-form. Take its dual, to get the 3-form $^*\boldsymbol{d}\phi$. Take its exterior derivative, to get the 4-form $\boldsymbol{d}^*\boldsymbol{d}\phi$. Take its dual, to get the scalar $\Box\phi \equiv -^*\boldsymbol{d}^*\boldsymbol{d}\phi$. Verify by index manipulations that $\Box$ as defined here is the wave operator; i.e., in any Lorentz frame, $\Box\phi = \phi_{,\alpha}{}^{,\alpha} = -(\partial^2\phi/\partial t^2) + \boldsymbol{\nabla}^2\phi$.

Start with 1-form $\boldsymbol{A}$. Get 2-form $\boldsymbol{F} = \boldsymbol{dA}$. Take its dual $^*\boldsymbol{F} = {}^*\boldsymbol{dA}$, also a 2-form. Take its exterior derivative, obtaining the 3-form $\boldsymbol{d}^*\boldsymbol{F}$ (has value $4\pi\, ^*\boldsymbol{J}$ in electromagnetism). Take its dual, obtaining the 1-form $^*\boldsymbol{d}^*\boldsymbol{F} = {}^*\boldsymbol{d}^*\boldsymbol{dA} = 4\pi\boldsymbol{J}$ ("Wave equation for electromagnetic 4-potential"). Reduce in index notation to

$$F_{\mu\nu}{}^{,\nu} = A_{\nu,\mu}{}^{,\nu} - A_{\mu,\nu}{}^{,\nu} = 4\pi J_\mu.$$

[More in Flanders (1963) or Misner and Wheeler (1957); see also exercise 3.17.]

§4.7. DISTANT ACTION FROM LOCAL LAW

Differential forms are a powerful tool in electromagnetic theory, but full power requires mastery of other tools as well. Action-at-a-distance techniques ("Green's functions," "propagators") are of special importance. Moreover, the passage from Maxwell field equations to electromagnetic action at a distance provides a preview of how Einstein's local equations will reproduce (approximately) Newton's $1/r^2$ law.

In flat spacetime and in a Lorentz coordinate system, express the coordinates of particle A as a function of its proper time α, thus:

$$a^\mu = a^\mu(\alpha), \qquad \frac{da^\mu}{d\alpha} = \dot{a}^\mu(\alpha), \qquad \frac{d^2a^\mu}{d\alpha^2} = \ddot{a}^\mu(\alpha). \tag{4.32}$$

Dirac found it helpful to express the distribution of charge and current for a particle of charge e following such a motion as a superposition of charges that momentarily

World line of charge regarded as succession of flash-on, flash-off charges

flash into existence and then flash out of existence. Any such flash has a localization in space and time that can be written as the product of four Dirac delta functions [see, for example, Schwartz (1950–1951), Lighthill (1958)]:

$$\delta^4(x^\mu - a^\mu) = \delta[x^0 - a^0(\alpha)]\,\delta[x^1 - a^1(\alpha)]\,\delta[x^2 - a^2(\alpha)]\,\delta[x^3 - a^3(\alpha)]. \quad (4.33)$$

Here any single Dirac function $\delta(x)$ ("symbolic function"; "distribution"; "limit of a Gauss error function" as width is made indefinitely narrow and peak indefinitely high, with integrated value always unity) both (1) vanishes for $x \neq 0$, and (2) has the integral $\int_{-\infty}^{+\infty} \delta(x)\,dx = 1$. Described in these terms, the density-current vector for the particle has the value ("superposition of flashes")

$$J^\mu = e \int \delta^4[x^\nu - a^\nu(\alpha)]\dot{a}^\mu(\alpha)\,d\alpha. \quad (4.34)$$

The density-current (4.34) drives the electromagnetic field, $\boldsymbol{F}$. Write $\boldsymbol{F} = \boldsymbol{dA}$ to satisfy automatically half of Maxwell's equations ($\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$):

$$F_{\mu\alpha} = \frac{\partial A_\alpha}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\alpha}. \quad (4.35)$$

In flat space, the remainder of Maxwell's equations ($\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\,{}^*\boldsymbol{J}$) become

$$\frac{\partial F_\mu{}^\nu}{\partial x^\nu} = 4\pi J_\mu$$

or

$$\frac{\partial}{\partial x^\mu}\frac{\partial A^\nu}{\partial x^\nu} - \eta^{\nu\alpha}\frac{\partial^2 A_\mu}{\partial x^\nu\,\partial x^\alpha} = 4\pi J_\mu. \quad (4.36)$$

Make use of the freedom that exists in the choice of 4-potentials A^ν to demand

$$\frac{\partial A^\nu}{\partial x^\nu} = 0 \quad (4.37)$$

(Lorentz gauge condition; see exercise 3.17). Thus get

The electromagnetic wave equation

$$\Box A_\mu = -4\pi J_\mu. \quad (4.38)$$

The density-current being the superposition of "flashes," the effect ($\boldsymbol{A}$) of this density-current can be expressed as the superposition of the effects E of elementary flashes; thus

The solution of the wave equation

$$A^\mu(x) = \int E[x - a(\alpha)]\dot{a}^\mu(\alpha)\,d\alpha, \quad (4.39)$$

where the "elementary effect" E ("kernel"; "Green's function") satisfies the equation

$$\Box E(x) = -4\pi\,\delta^4(x). \quad (4.40)$$

One solution is the "half-advanced-plus-half-retarded potential,"

$$E(x) = \delta(\eta_{\alpha\beta}x^\alpha x^\beta). \quad (4.41)$$

It vanishes everywhere except on the backward and forward light cones, where it has equal strength. Normally more useful is the retarded solution,

$$R(x) = \begin{cases} 2E(x) & \text{if } x^0 > 0, \\ 0 & \text{if } x^0 < 0, \end{cases} \tag{4.42}$$

which is obtained by doubling (4.41) in the region of the forward light cone and nullifying it in the region of the backward light cone. All electrodynamics (Coulomb forces, Ampère's law, electromagnetic induction, radiation) follows from the simple expression (4.39) for the vector potential [see, e.g., Wheeler and Feynman (1945) and (1949), also Rohrlich (1965)].

EXERCISES

Exercise 4.1. GENERIC LOCAL ELECTROMAGNETIC FIELD EXPRESSED IN SIMPLEST FORM

In the laboratory Lorentz frame, the electric field is $\boldsymbol{E}$, the magnetic field $\boldsymbol{B}$. Special cases are: (1) pure electric field ($\boldsymbol{B} = 0$); (2) pure magnetic field ($\boldsymbol{E} = 0$); and (3) "radiation field" or "null field" ($\boldsymbol{E}$ and $\boldsymbol{B}$ equal in magnitude and perpendicular in direction). All cases other than (1), (2), and (3) are "generic." In the generic case, calculate the Poynting density of flow of energy $\boldsymbol{E} \times \boldsymbol{B}/4\pi$ and the density of energy $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$. Define the direction of a unit vector $\boldsymbol{n}$ and the magnitude of a velocity parameter α by the ratio of energy flow to energy density:

$$\boldsymbol{n} \tanh 2\alpha = \frac{2\boldsymbol{E} \times \boldsymbol{B}}{\boldsymbol{E}^2 + \boldsymbol{B}^2}.$$

View the same electromagnetic field in a rocket frame moving in the direction of $\boldsymbol{n}$ with the velocity parameter α (not 2α; factor 2 comes in because energy flow and energy density are components, not of a vector, but of a tensor). By employing the formulas for a Lorentz transformation (equation 3.23), or otherwise, show that the energy flux vanishes in the rocket frame, with the consequence that $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$ are parallel. No one can prevent the $\bar{z}$-axis from being put in the direction common to $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$. Show that with this choice of direction, ***Faraday*** becomes

$$\boldsymbol{F} = \bar{E}_z\, \boldsymbol{d}\bar{z} \wedge \boldsymbol{d}\bar{t} + \bar{B}_z\, \boldsymbol{d}\bar{x} \wedge \boldsymbol{d}\bar{y}$$

(only two wedge products needed to represent the generic local field; "canonical representation"; valid in one frame, valid in any frame).

Exercise 4.2. FREEDOM OF CHOICE OF 1-FORMS IN CANONICAL REPRESENTATION OF GENERIC LOCAL FIELD

Deal with a region so small that the variation of the field from place to place can be neglected. Write ***Faraday*** in canonical representation in the form

$$\boldsymbol{F} = \boldsymbol{d}p_I \wedge \boldsymbol{d}q^I + \boldsymbol{d}p_{II} \wedge \boldsymbol{d}q^{II},$$

where p_A ($A = I$ or II) and q^A are scalar functions of position in spacetime. Define a "canonical transformation" to new scalar functions of position $p_{\bar{A}}$ and $q^{\bar{A}}$ by way of the "equation of transformation"

$$p_A\, \boldsymbol{d}q^A = \boldsymbol{d}S + p_{\bar{A}}\, \boldsymbol{d}q^{\bar{A}},$$

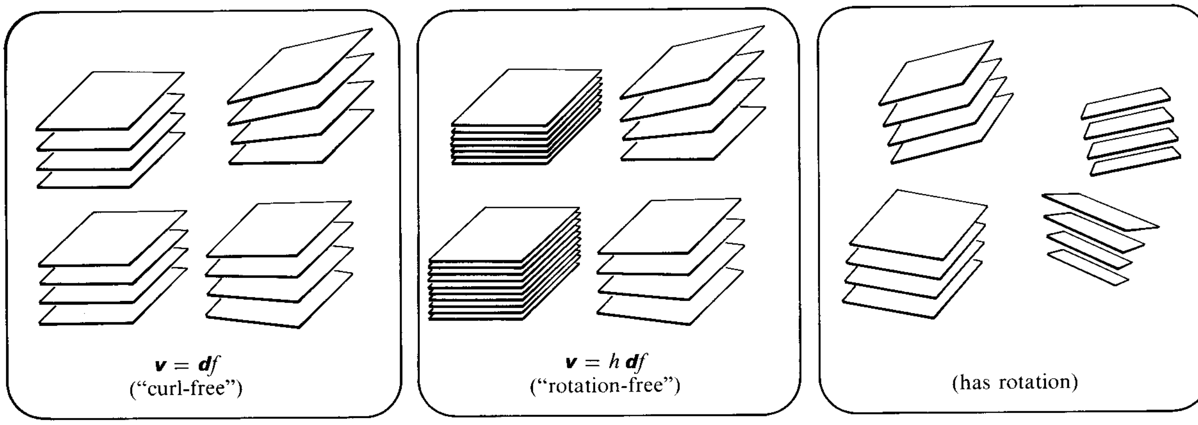


Figure 4.7.
Some simple types of 1-forms compared and contrasted.

where the "generating function" S of the transformation is an arbitrary function of the q^A and the $q^{\bar{A}}$:

$$\boldsymbol{d}S = (\partial S/\partial q^A)\,\boldsymbol{d}q^A + (\partial S/\partial q^{\bar{A}})\,\boldsymbol{d}q^{\bar{A}}.$$

(a) Derive expressions for the two p_A's and the two $p_{\bar{A}}$'s in terms of S by equating coefficients of dq^I, dq^{II}, $dq^{\bar{I}}$, $dq^{\bar{II}}$ individually on the two sides of the equation of transformation.

(b) Use these expressions for the p_A's and $p_{\bar{A}}$'s to show that $\boldsymbol{F} = \boldsymbol{d}p_A \wedge \boldsymbol{d}q^A$ and $\bar{\boldsymbol{F}} = \boldsymbol{d}p_{\bar{A}} \wedge \boldsymbol{d}q^{\bar{A}}$, ostensibly different, are actually expressions for one and the same 2-form in terms of alternative sets of 1-forms.

Exercise 4.3. A CLOSED OR CURL-FREE 1-FORM IS A GRADIENT

Given a 1-form $\boldsymbol{\sigma}$ such that $\boldsymbol{d\sigma} = 0$, show that $\boldsymbol{\sigma}$ can be expressed in the form $\boldsymbol{\sigma} = \boldsymbol{d}f$, where f is some scalar. The 1-form $\boldsymbol{\sigma}$ is said to be "curl-free," a narrower category of 1-form than the "rotation-free" 1-form of the next exercise (expressible as $\boldsymbol{\sigma} = h\,\boldsymbol{d}f$), and it in turn is narrower (see Figure 4.7) than the category of "1-forms with rotation" (not expressible in the form $\boldsymbol{\sigma} = h\,\boldsymbol{d}f$). When the 1-form $\boldsymbol{\sigma}$ is expressed in terms of basis 1-forms $\boldsymbol{d}x^\alpha$, multiplied by corresponding components σ_α, show that "curl-free" implies $\sigma_{[\alpha,\beta]} = 0$.

Exercise 4.4. CANONICAL EXPRESSION FOR A ROTATION-FREE 1-FORM

In three dimensions a rigid body turning with angular velocity ω about the z-axis has components of velocity $v_y = \omega x$, and $v_x = -\omega y$. The quantity curl $\boldsymbol{v} = \boldsymbol{\nabla} \times \boldsymbol{v}$ has z-component equal to 2ω, and all other components equal to zero. Thus the scalar product of $\boldsymbol{v}$ and curl $\boldsymbol{v}$ vanishes:

$$v_{[i,j}v_{k]} = 0.$$

The same concept generalizes to four dimensions,

$$v_{[\alpha,\beta}v_{\gamma]} = 0,$$

and lends itself to expression in coordinate-free language, as the requirement that a certain 3-form must vanish:

$$\boldsymbol{d}\boldsymbol{v} \wedge \boldsymbol{v} = 0.$$

Any 1-form $\boldsymbol{v}$ satisfying this condition is said to be "rotation-free." Show that a 1-form is rotation-free if and only if it can be written in the form

$$\boldsymbol{v} = h\, \boldsymbol{d}f,$$

where h and f are scalar functions of position (the "Frobenius theorem").

Exercise 4.5. FORMS ENDOWED WITH POLAR SINGULARITIES

List the principal results on how such forms are representable, such as

$$\boldsymbol{\Phi}_1 = \frac{\boldsymbol{d}S}{S} \wedge \boldsymbol{\psi}_1 + \boldsymbol{\theta}_1,$$

and the conditions under which each applies [for the meaning and answer to this exercise, see Lascoux (1968)].

Exercise 4.6. THE FIELD OF THE OSCILLATING DIPOLE

Verify that the expressions given for the electromagnetic field of an oscillating dipole in equations (4.23) and (4.24) satisfy $\boldsymbol{dF} = 0$ everywhere and $\boldsymbol{d}{}^*\boldsymbol{F} = 0$ everywhere except at the origin.

Exercise 4.7. THE 2-FORM MACHINERY TRANSLATED INTO TENSOR MACHINERY

This exercise is stated at the end of the legend caption of Figure 4.1.

Exercise 4.8. PANCAKING THE COULOMB FIELD

Figure 4.5 shows a spacelike slice, $t =$ const, through the ***Maxwell*** of a point-charge at rest. By the following pictorial steps, verify that the electric-field lines get compressed into the transverse direction when viewed from a moving Lorentz frame: (1) Draw a picture of an equatorial slice ($\theta = \pi/2$; t, r, ϕ variable) through ***Maxwell*** $= {}^*\boldsymbol{F}$. (2) Draw various spacelike slices, corresponding to constant time in various Lorentz frames, through the resultant geometric structure. (3) Interpret the intersection of ***Maxwell*** $= {}^*\boldsymbol{F}$ with each Lorentz slice in the manner of Figure 4.3.

Exercise 4.9. COMPUTATION OF SURFACE INTEGRALS

In Box 4.1 the definition

$$\int \boldsymbol{\alpha} = \int \cdots \int \left\langle \boldsymbol{\alpha}, \frac{\partial \mathscr{P}}{\partial \lambda^1} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial \lambda^p} \right\rangle d\lambda^1 \ldots d\lambda^p$$

is given for the integral of a p-form $\boldsymbol{\alpha}$ over a p-surface $\mathscr{P}(\lambda^1, \ldots, \lambda^p)$ in n-dimensional space. From this show that the following computational rule (also given in Box 4.1) works: (1) substitute the equation for the surface,

$$x^k = x^k(\lambda^1, \ldots, \lambda^p),$$

into $\boldsymbol{\alpha}$ and collect terms in the form

$$\boldsymbol{\alpha} = a(\lambda^1, \ldots, \lambda^p)\, \boldsymbol{d}\lambda^1 \wedge \cdots \wedge \boldsymbol{d}\lambda^p;$$

(2) integrate

$$\int \boldsymbol{\alpha} = \int \cdots \int a(\lambda^1, \ldots, \lambda^p)\, d\lambda^1 \ldots d\lambda^p$$

using the elementary definition of integration.

Exercise 4.10. WHITAKER'S CALUMOID, OR, THE LIFE OF A LOOP

Take a closed loop, bounding a 2-dimensional surface $\mathcal{S}$. It entraps a certain flux of ***Faraday*** $\boldsymbol{\Phi}_{F} = \int_{\mathcal{S}} \boldsymbol{F}$ ("magnetic tubes") and a certain flux of ***Maxwell*** $\Phi_{M} = \int_{\mathcal{S}} {}^{*}\boldsymbol{F}$ ("electric tubes").

(a) Show that the fluxes Φ_F and Φ_M depend only on the choice of loop, and not on the choice of the surface $\mathcal{S}$ bounded by the loop, if and only if $\boldsymbol{dF} = \boldsymbol{d}{}^{*}\boldsymbol{F} = 0$ (no magnetic charge; no electric charge). *Hint:* use generalized Stokes theorem, Boxes 4.1 and 4.6.

(b) Move the loop in space and time so that it continues to entrap the same two fluxes. Move it forward a little more here, a little less there, so that it continues to do so. In this way trace out a 2-dimensional surface ("calumoid"; see E. T. Whitaker 1904) $\mathcal{P} = \mathcal{P}(a, b)$; $x^{\mu} = x^{\mu}(a, b)$. Show that the elementary bivector in this surface, $\boldsymbol{\Sigma} = \partial\mathcal{P}/\partial a \wedge \partial\mathcal{P}/\partial b$ satisfies $\langle \boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ and $\langle {}^{*}\boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$.

(c) Show that these differential equations for $x^{\mu}(a, b)$ can possess a solution, with given initial condition $x^{\mu} = x^{\mu}(a, 0)$ for the initial location of the loop, if $\boldsymbol{dF} = 0$ and $\boldsymbol{d}{}^{*}\boldsymbol{F} = 0$ (no magnetic charge, no electric charge).

(d) Consider a static, uniform electric field $\boldsymbol{F} = -E_x\, \boldsymbol{dt} \wedge \boldsymbol{dx}$. Solve the equations, $\langle \boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ and $\langle {}^{*}\boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ to find the equation $\mathcal{P}(a, b)$ for the most general calumoid. [*Answer:* $y = y(a)$, $z = z(a)$, $x = x(b)$, $t = t(b)$.] Exhibit two special cases: (i) a calumoid that lies entirely in a hypersurface of constant time [loop moves at infinite velocity; analogous to super-light velocity of point of crossing for two blades of a pair of scissors]; (ii) a calumoid whose loop remains forever at rest in the t, x, y, z Lorentz frame.

Exercise 4.11. DIFFERENTIAL FORMS AND HAMILTONIAN MECHANICS

Consider a dynamic system endowed with two degrees of freedom. For the definition of this system as a Hamiltonian system (special case: here the Hamiltonian is independent of time), one needs (1) a definition of canonical variables (see Box 4.5) and (2) a knowledge of the Hamiltonian H as a function of the coordinates q^1, q^2 and the canonically conjugate momenta p_1, p_2. To derive the laws of mechanics, consider the five-dimensional space of p_1, p_2, q^1, q^2, and t, and a curve in this space leading from starting values of the five coordinates (subscript A) to final values (subscript B), and the value

$$I = \int_A^B p_1\, \boldsymbol{d}q^1 + p_2\, \boldsymbol{d}q^2 - H(p, q)\, \boldsymbol{dt} = \int_A^B \boldsymbol{\omega}$$

of the integral I taken along this path. The difference of the integral for two "neighboring" paths enclosing a two-dimensional region $\mathcal{S}$, according to the theorem of Stokes (Boxes 4.1 and 4.6), is

$$\delta I = \oint \boldsymbol{\omega} = \int_{\mathcal{S}} \boldsymbol{d\omega}.$$

The principle of least action (principle of "extremal history") states that the representative point of the system must travel along a route in the five-dimensional manifold (route with tangent vector $d\mathcal{P}/dt$) such that the variation vanishes for this path; i.e.,

$$\boldsymbol{d\omega}(\ldots, d\mathcal{P}/dt) = 0$$

(2-form $\boldsymbol{d\omega}$ with a single vector argument supplied, and other slot left unfilled, gives the 1-form in 5-space that must vanish). This fixes only the direction of $d\mathcal{P}/dt$; its magnitude can be normalized by requiring $\langle \boldsymbol{dt}, d\mathcal{P}/dt \rangle = 1$.

(a) Evaluate $\boldsymbol{d\omega}$ from the expression $\boldsymbol{\omega} = p_j\, \boldsymbol{d}q^j - H\, \boldsymbol{dt}$.

(b) Set $d\mathcal{P}/dt = \dot{q}^j(\partial\mathcal{P}/\partial q^j) + \dot{p}_j(\partial\mathcal{P}/\partial p_j) + \dot{t}(\partial\mathcal{P}/\partial t)$, and expand $\boldsymbol{d\omega}(\ldots, d\mathcal{P}/dt) = 0$ in terms of the basis $\{\boldsymbol{d}p_j, \boldsymbol{d}q^k, \boldsymbol{dt}\}$.

Box 4.5 METRIC STRUCTURE AND HAMILTONIAN OR "SYMPLECTIC STRUCTURE" COMPARED AND CONTRASTED

	Metric structure	*Symplectic structure*
1. Physical application	Geometry of spacetime	Hamiltonian mechanics
2. Canonical structure	$(\ldots\cdot\ldots) = \text{“}\boldsymbol{ds}^2\text{”} = -\boldsymbol{dt}\otimes\boldsymbol{dt} + \boldsymbol{dx}\otimes\boldsymbol{dx} + \boldsymbol{dy}\otimes\boldsymbol{dy} + \boldsymbol{dz}\otimes\boldsymbol{dz}$	$\boldsymbol{\Theta} = \boldsymbol{d}p_1 \wedge \boldsymbol{d}q^1 + \boldsymbol{d}p_2 \wedge \boldsymbol{d}q^2$
3. Nature of "metric"	Symmetric	Antisymmetric
4. Name for given coordinate system and any other set of four coordinates in which metric has same form	Lorentz coordinate system	System of "canonically" (or "dynamically") conjugate coordinates
5. Field equation for this metric	$\boldsymbol{R}_{\mu\nu\alpha\beta} = 0$ (zero Riemann curvature; flat spacetime)	$\boldsymbol{d\Theta} = 0$ ("closed 2-form"; condition automatically satisfied by expression above).
6. The four-dimensional manifold	Spacetime	Phase space
7. Coordinate-free description of the structure of this manifold	$\textbf{Riemann} = 0$	$\boldsymbol{d\Theta} = 0$
8. Canonical coordinates distinguished from other coordinates (allowable but less simple)	Make metric take above form (item 2)	Make metric take above form (item 2)

(c) Show that this five-dimensional equation can be written in the 4-dimensional phase space of $\{q^j, p_k\}$ as

$$\boldsymbol{\Theta}(\ldots, d\mathscr{P}/dt) = \boldsymbol{d}H,$$

where $\boldsymbol{\Theta}$ is the 2-form defined in Box 4.5.

(d) Show that the components of $\boldsymbol{\Theta}(\ldots, d\mathscr{P}/dt) = \boldsymbol{d}H$ in the $\{q^j, p_k\}$ coordinate system are the familiar Hamilton equations. Note that this conclusion depends only on the form assumed for $\boldsymbol{\Theta}$, so that one also obtains the standard Hamilton equations in any other phase-space coordinates $\{\bar{q}^j, \bar{p}_k\}$ ("canonical variables") for which

$$\boldsymbol{\Theta} = \boldsymbol{d}\bar{p}_1 \wedge \boldsymbol{d}\bar{q}^1 + \boldsymbol{d}\bar{p}_2 \wedge \boldsymbol{d}\bar{q}^2.$$

Exercise 4.12. SYMMETRY OPERATIONS AS TENSORS

We define the meaning of square and round brackets enclosing a set of indices as follows:

$$V_{(\alpha_1\ldots\alpha_p)} \equiv \frac{1}{p!}\Sigma V_{\alpha_{\pi_1}\ldots\alpha_{\pi_p}}, \qquad V_{[\alpha_1\ldots\alpha_p]} \equiv \frac{1}{p!}\Sigma(-1)^{\pi} V_{\alpha_{\pi_1}\ldots\alpha_{\pi_p}}.$$

Box 4.6 BIRTH OF STOKES' THEOREM

Central to the mathematical formulation of electromagnetism are the theorems of Gauss (taken up in Chapter 5) and Stokes. Both today appear together as one unity when expressed in the language of forms. In earlier times the unity was not evident. Everitt (1970) recalls the history of Stokes' theorem: "The Smith's Prize paper set by [G. C.] Stokes [Lucasian Professor of Mathematics] and taken by Maxwell in [February] 1854 . . .

> 5. Given the centre and two points of an ellipse, and the length of the major axis, find its direction by a geometrical construction.
>
> 6. Integrate the differential equation
>
> $$(a^2 - x^2)\, dy^2 + 2xydydx + (a^2 - y^2)\, dx^2 = 0.$$
>
> Has it a singular solution?
>
> 7. In a double system of curves of double curvature, a tangent is always drawn at the variable point P; shew that, as P moves away from an arbitrary fixed point Q, it must begin to move along a generating line of an elliptic cone having Q for vertex in order that consecutive tangents may ultimately intersect, but that the conditions of the problem may be impossible.
>
> 8. If X, Y, Z be functions of the rectangular co-ordinates x, y, z, dS an element of any limited surface, l, m, n the cosines of the inclinations of the normal at dS to the axes, ds an element of the bounding line, shew that
>
> $$\iint \left\{ l\left(\frac{dZ}{dy} - \frac{dY}{dz}\right) + m\left(\frac{dX}{dz} - \frac{dZ}{dx}\right) + n\left(\frac{dY}{dx} - \frac{dX}{dy}\right)\right\} dS$$
> $$= \int \left(X\frac{dx}{ds} + Y\frac{dy}{ds} + Z\frac{dz}{ds}\right) ds,$$
>
> the differential coefficients of X, Y, Z being partial, and the single integral being taken all round the perimeter of the surface

marks the first appearance in print of the formula connecting line and surface integrals now known as Stokes' theorem. This was of great importance to Maxwell's development of electromagnetic theory. The earliest explicit proof of the theorem appears to be that given in a letter from Thomson to Stokes dated July 2, 1850." [Quoted in Campbell and Garnett (1882), pp. 186–187.]

Here the sum is taken over all permutations π of the numbers $1, 2, \ldots, p$, and $(-1)^\pi$ is $+1$ or -1 depending on whether the permutation is even or odd. The quantity V may have other indices, not shown here, besides the set of p indices $\alpha_1, \alpha_2, \ldots, \alpha_p$, but only this set of indices is affected by the operations described here. The numbers $\pi_1, \pi_2, \ldots, \pi_p$ are the numbers $1, 2, \ldots, p$ rearranged according to the permutation π. (Cases $p = 2, 3$ were treated in exercise 3.12.) We therefore have machinery to convert any rank-p tensor with components $V_{\alpha_1\ldots\alpha_p}$ into a new tensor with components

$$[\boldsymbol{Alt}(\boldsymbol{V})]_{\mu_1\ldots\mu_p} = V_{[\mu_1\ldots\mu_p]}.$$

Since this machinery $\boldsymbol{Alt}$ is linear, it can be viewed as a tensor which, given suitable arguments $\boldsymbol{u}, \boldsymbol{v}, \ldots, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ produces a number

$$u^\mu v^\nu \ldots w^\lambda \alpha_{[\mu}\beta_\nu \ldots \gamma_{\lambda]}.$$

(a) Show that the components of this tensor are

$$(\boldsymbol{Alt})_{\beta_1\ldots\beta_p}{}^{\alpha_1\ldots\alpha_p} = (p!)^{-1}\, \delta^{\alpha_1\ldots\alpha_p}_{\beta_1\ldots\beta_p}$$

(Note: indices of δ are almost never raised or lowered, so this notation leads to no confusion.)

where

$$\delta^{\alpha_1\ldots\alpha_p}_{\beta_1\ldots\beta_p} = \begin{cases} +1 \text{ if } (\alpha_1, \ldots, \alpha_p) \text{ is an even permutation of } (\beta_1, \ldots, \beta_p), \\ -1 \text{ if } (\alpha_1, \ldots, \alpha_p) \text{ is an odd permutation of } (\beta_1, \ldots, \beta_p), \\ 0 \text{ if (i) any two of the } \alpha\text{'s are the same,} \\ 0 \text{ if (ii) any two of the } \beta\text{'s are the same,} \\ 0 \text{ if (iii) the } \alpha\text{'s and } \beta\text{'s are different sets of integers.} \end{cases}$$

Note that the demonstration, and therefore these component values, are correct in any frame.

(b) Show for any "alternating" (i.e., "*completely antisymmetric*") tensor $A_{\alpha_1\ldots\alpha_p} = A_{[\alpha_1\ldots\alpha_p]}$ that

$$\begin{aligned} \frac{1}{p!} A_{\alpha_1\ldots\alpha_p}\, \delta^{\alpha_1\ldots\alpha_p\beta_1\ldots\beta_q}_{\gamma_1\ldots\gamma_p\gamma_{p+1}\ldots\gamma_{p+q}} &= \sum_{\alpha_1<\alpha_2<\ldots<\alpha_p} A_{\alpha_1\ldots\alpha_p}\, \delta^{\alpha_1\ldots\alpha_p\beta_1\ldots\beta_q}_{\gamma_1\ldots\ldots\ldots\gamma_{p+q}} \\ &\equiv A_{|\alpha_1\ldots\alpha_p|}\, \delta^{\alpha_1\ldots\alpha_p\beta_1\ldots\beta_q}_{\gamma_1\ldots\ldots\ldots\gamma_{p+q}}. \end{aligned}$$

The final line here introduces the convention that a summation over indices enclosed between vertical bars includes only terms with those indices in increasing order. Show, consequently or similarly, that

$$\delta^{\alpha_1\ldots\alpha_p\beta_1\ldots\beta_q}_{\gamma_1\ldots\ldots\ldots\gamma_{p+q}}\, \delta^{\mu_1\ldots\mu_q}_{|\beta_1\ldots\beta_q|} = \delta^{\alpha_1\ldots\alpha_p\mu_1\ldots\mu_q}_{\gamma_1\ldots\ldots\ldots\gamma_{p+q}}.$$

(c) *Define* the exterior ("wedge") product of any two alternating tensors by

$$(\boldsymbol{\alpha} \wedge \boldsymbol{\beta})_{\lambda_1\ldots\lambda_{p+q}} = \delta^{\mu_1\ldots\mu_p\,\nu_1\ldots\nu_q}_{\lambda_1\ldots\lambda_p\lambda_{p+1}\ldots\lambda_{p+q}}\, \alpha_{|\mu_1\ldots\mu_p|}\, \beta_{|\nu_1\ldots\nu_q|};$$

and similarly

$$(\boldsymbol{U} \wedge \boldsymbol{V})^{\lambda_1\ldots\lambda_{p+q}} = \delta^{\lambda_1\ldots\lambda_p\lambda_{p+1}\ldots\lambda_{p+q}}_{\mu_1\ldots\mu_p\nu_1\ldots\nu_q}\, U^{|\mu_1\ldots\mu_p|}\, V^{|\nu_1\ldots\nu_q|}.$$

Show that this implies equation (3.45b). Establish the associative law for this product rule by showing that

$$
\begin{aligned}
&[(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) \wedge \boldsymbol{\gamma}]_{\sigma_1 \ldots \sigma_{p+q+r}} \\
&= \delta^{\lambda_1 \ldots \lambda_p \mu_1 \ldots \mu_q \nu_1 \ldots \nu_r}_{\sigma_1 \ldots \sigma_{p+q+r}} \alpha_{|\lambda_1 \ldots \lambda_p|} \beta_{|\mu_1 \ldots \mu_q|} \gamma_{|\nu_1 \ldots \nu_r|} \\
&= [\boldsymbol{\alpha} \wedge (\boldsymbol{\beta} \wedge \boldsymbol{\gamma})]_{\sigma_1 \ldots \sigma_{p+q+r}};
\end{aligned}
$$

and show that this reduces to the 3-form version of Equation (3.45c) when $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are all 1-forms.

(d) Derive the following formula for the components of the exterior product of p vectors

$$
\begin{aligned}
(\boldsymbol{u}_1 \wedge \boldsymbol{u}_2 \wedge \cdots \wedge \boldsymbol{u}_p)^{\alpha_1 \ldots \alpha_p} &= \delta^{\alpha_1 \ldots \alpha_p}_{\mu \ldots \nu} (u_1)^\mu \ldots (u_p)^\nu \\
&= p! \, u_1{}^{[\alpha_1} u_2{}^{\alpha_2} \ldots u_p{}^{\alpha_p]} \\
&= \delta^{\alpha_1 \alpha_2 \ldots \alpha_p}_{1\ 2\ \ldots\ p} \det [(u_\mu)^\lambda].
\end{aligned}
$$

CHAPTER 5

STRESS-ENERGY TENSOR AND CONSERVATION LAWS

§5.1. TRACK-1 OVERVIEW

"Geometry tells matter how to move, and matter tells geometry how to curve." However, it will do no good to look into curvature (Part III) and Einstein's law for the production of curvature by mass-energy (Part IV) until a tool can be found to determine how much mass-energy there is in a unit volume. That tool is the stress-energy tensor. It is the focus of attention in this chapter.

The essential features of the stress-energy tensor are summarized in Box 5.1 for the benefit of readers who want to rush on into gravitation physics as quickly as possible. Such readers can proceed directly from Box 5.1 into Chapter 6—though by doing so, they close the door on several later portions of track two, which lean heavily on material treated in this chapter.

The rest of this chapter is Track 2.
It depends on no preceding Track-2 material.
It is needed as preparation for Chapter 20 (conservation laws for mass and angular momentum).
It will be extremely helpful in all applications of gravitation theory (Chapters 18–40).

§5.2. THREE-DIMENSIONAL VOLUMES AND DEFINITION OF THE STRESS-ENERGY TENSOR

Spacetime contains a flowing "river" of 4-momentum. Each particle carries its 4-momentum vector with itself along its world line. Many particles, on many world lines, viewed in a smeared-out manner (continuum approximation), produce a continuum flow—a river of 4-momentum. Electromagnetic fields, neutrino fields, meson fields: they too contribute to the river.

How can the flow of the river be quantified? By means of a linear machine: the stress-energy tensor $\boldsymbol{T}$.

Choose a small, three-dimensional parallelepiped in spacetime with vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ for edges (Figure 5.1). Ask how much 4-momentum crosses that volume in

Box 5.1 CHAPTER 5 SUMMARIZED

A. STRESS-ENERGY TENSOR AS A MACHINE

At each event in spacetime, there exists a stress-energy tensor. It is a machine that contains a knowledge of the energy density, momentum density, and stress as measured by any and all observers at that event. Included are energy, momentum, and stress associated with all forms of matter and all nongravitational fields.

The stress-energy tensor is a linear, symmetric machine with two slots for the insertion of two vectors: $\boldsymbol{T}(\ldots, \ldots)$. Its output, for given input, can be summarized as follows.

(1) Insert the 4-velocity $\boldsymbol{u}$ of an observer into one of the slots; leave the other slot empty. The output is

$$\boldsymbol{T}(\boldsymbol{u}, \ldots) = \boldsymbol{T}(\ldots, \boldsymbol{u}) = -\begin{pmatrix}\text{density of 4-momentum,} \\ \text{“}d\boldsymbol{p}/dV\text{,” i.e., 4-momentum} \\ \text{per unit of three-dimensional volume,} \\ \text{as measured in observer's} \\ \text{Lorentz frame at event where} \\ \boldsymbol{T}\text{ is chosen}\end{pmatrix};$$

i.e., $T^{\alpha}{}_{\beta}u^{\beta} = T_{\beta}{}^{\alpha}u^{\beta} = -(dp^{\alpha}/dV)$ for observer with 4-velocity u^{α}.

(2) Insert 4-velocity of observer into one slot; insert an arbitrary unit vector $\boldsymbol{n}$ into the other slot. The output is

$$\boldsymbol{T}(\boldsymbol{u}, \boldsymbol{n}) = \boldsymbol{T}(\boldsymbol{n}, \boldsymbol{u}) = -\begin{pmatrix}\text{component, “}\boldsymbol{n}\cdot d\boldsymbol{p}/dV\text{”, of} \\ \text{4-momentum density along the} \\ \boldsymbol{n}\text{ direction, as measured in} \\ \text{observer's Lorentz frame}\end{pmatrix};$$

i.e., $T_{\alpha\beta}u^{\alpha}n^{\beta} = T_{\alpha\beta}n^{\alpha}u^{\beta} = -n_{\mu}dp^{\mu}/dV$.

(3) Insert 4-velocity of observer into both slots. The output is the density of mass-energy that he measures in his Lorentz frame:

$$\boldsymbol{T}(\boldsymbol{u}, \boldsymbol{u}) = \begin{pmatrix}\text{mass-energy per unit volume as measured} \\ \text{in frame with 4-velocity } \boldsymbol{u}\end{pmatrix}.$$

(4) Pick an observer and choose two spacelike basis vectors, $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$, of his Lorentz frame. Insert $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$ into the slots of $\boldsymbol{T}$. The output is the j,k component of the stress as measured by that observer:

$$\begin{aligned} T_{jk} &= \boldsymbol{T}(\boldsymbol{e}_j, \boldsymbol{e}_k) = T_{kj} = \boldsymbol{T}(\boldsymbol{e}_k, \boldsymbol{e}_j) \\ &= \begin{pmatrix}j\text{-component of force acting} \\ \text{from side } x^k - \varepsilon \text{ to side } x^k + \varepsilon\text{,} \\ \text{across a unit surface area with} \\ \text{perpendicular direction } \boldsymbol{e}_k\end{pmatrix} = \begin{pmatrix}k\text{-component of force acting} \\ \text{from side } x^j - \varepsilon \text{ to side } x^j + \varepsilon\text{,} \\ \text{across a unit surface area with} \\ \text{perpendicular direction } \boldsymbol{e}_j\end{pmatrix}. \end{aligned}$$

Box 5.1 ***(continued)***

B. STRESS-ENERGY TENSOR FOR A PERFECT FLUID

One type of matter studied extensively later in this book is a *"perfect fluid."* A perfect fluid is a fluid or gas that (1) moves through spacetime with a 4-velocity $\boldsymbol{u}$ which may vary from event to event, and (2) exhibits a density of mass-energy ρ and an isotropic pressure p in the rest frame of each fluid element. Shear stresses, anisotropic pressures, and viscosity must be absent, or the fluid is not perfect. The stress-energy tensor for a perfect fluid at a given event can be constructed from the metric tensor, $\boldsymbol{g}$, the 4-velocity, $\boldsymbol{u}$, and the rest-frame density and pressure, ρ and p:

$$\boldsymbol{T} = (\rho + p)\boldsymbol{u} \otimes \boldsymbol{u} + p\boldsymbol{g}, \qquad \text{or } T_{\alpha\beta} = (\rho + p)u_\alpha u_\beta + pg_{\alpha\beta}.$$

In the fluid's rest frame, the components of this stress-energy tensor have the expected form (insert into a slot of $\boldsymbol{T}$, as 4-velocity of observer, just the fluid's 4-velocity):

$$T^\alpha{}_\beta u^\beta = [(\rho + p)u^\alpha u_\beta + p\delta^\alpha{}_\beta]u^\beta = -(\rho + p)u^\alpha + pu^\alpha = -\rho u^\alpha;$$

i.e.,

$$T^0{}_\beta u^\beta = -\rho = -(\text{mass-energy density}) = -dp^0/dV,$$
$$T^j{}_\beta u^\beta = 0 = -(\text{momentum density}) = -dp^j/dV;$$

also

$$T_{jk} = \boldsymbol{T}(\boldsymbol{e}_j, \boldsymbol{e}_k) = p\delta_{jk} = \text{stress-tensor components}.$$

C. CONSERVATION OF ENERGY-MOMENTUM

In electrodynamics the conservation of charge can be expressed by the differential equation

$$\partial(\text{charge density})/\partial t + \boldsymbol{\nabla} \cdot (\text{current density}) = 0;$$

i.e., $J^0{}_{,0} + \boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$; i.e. $J^\alpha{}_{,\alpha} = 0$; i.e., $\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$. Similarly, conservation of energy-momentum can be expressed by the fundamental geometric law

$$\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0.$$

(Because $\boldsymbol{T}$ is symmetric, it does not matter on which slot the divergence is taken.) This law plays an important role in gravitation theory.

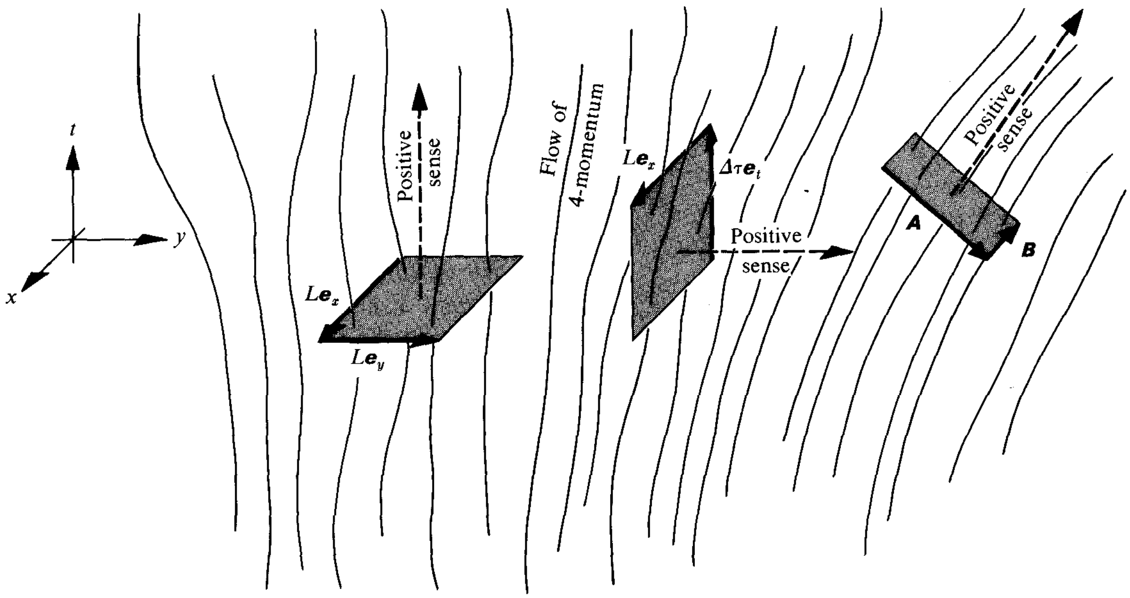


Figure 5.1.
The "river" of 4-momentum flowing through spacetime, and three different 3-volumes across which it flows. (One dimension is suppressed from the picture; so the 3-volumes look like 2-volumes.) The first 3-volume is the interior of a cubical soap box momentarily at rest in the depicted Lorentz frame. Its edges are $L\boldsymbol{e}_x$, $L\boldsymbol{e}_y$, $L\boldsymbol{e}_z$; and its volume 1-form, with "positive" sense toward future ("standard orientation"), is $\boldsymbol{\Sigma} = L^3\,\boldsymbol{d}t = -V\boldsymbol{u}$($V = L^3$ = volume as measured in rest frame; $\boldsymbol{u} = -\boldsymbol{d}t$ = 4-velocity of box). The second 3-volume is the "world sheet" swept out in time $\Delta\tau$ by the top of a second cubical box. The box top's edges are $L\boldsymbol{e}_x$ and $L\boldsymbol{e}_z$; and its volume 1-form, with "positive" sense away from the box's interior, in direction of increasing y, is $\boldsymbol{\Sigma} = L^2\,\Delta\tau\,\boldsymbol{d}y = \mathcal{A}\,\Delta\tau\boldsymbol{\sigma}$ ($\mathcal{A} = L^2$ = area of box top; $\boldsymbol{\sigma} = \boldsymbol{d}y$ = unit 1-form containing world tube). The third 3-volume is an arbitrary one, with edges $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and volume 1-form $\Sigma_\mu = \epsilon_{\mu\alpha\beta\gamma}A^\alpha B^\beta C^\gamma$.

its positive sense (i.e., from its "negative side" toward its "positive side"). To calculate the answer: (1) Construct the *"volume 1-form"*

Mathematical representation of 3-volumes

$$\Sigma_\mu = +\epsilon_{\mu\alpha\beta\gamma}A^\alpha B^\beta C^\gamma; \tag{5.1}$$

the parallelepiped lies in one of the 1-form surfaces, and the positive sense across the parallelepiped is defined to be the positive sense of the 1-form $\boldsymbol{\Sigma}$. (2) Insert this volume 1-form into the second slot of the stress-energy tensor $\boldsymbol{T}$. The result is

Momentum crossing a 3-volume calculated, using stress-energy tensor

$$\underset{\text{empty slot}}{\boldsymbol{T}(\underbrace{\ldots}},\boldsymbol{\Sigma}) = \boldsymbol{p} = \begin{pmatrix}\text{momentum crossing from}\\ \text{negative side toward positive side}\end{pmatrix}. \tag{5.2}$$

(3) To get the projection of the 4-momentum along a vector $\boldsymbol{w}$ or 1-form $\boldsymbol{\alpha}$, insert the volume 1-form $\boldsymbol{\Sigma}$ into the second slot and $\boldsymbol{w}$ or $\boldsymbol{\alpha}$ into the first:

$$\boldsymbol{T}(\boldsymbol{w},\boldsymbol{\Sigma}) = \boldsymbol{w}\cdot\boldsymbol{p}, \qquad \boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\Sigma}) = \langle\boldsymbol{\alpha},\boldsymbol{p}\rangle. \tag{5.3}$$

This defines the stress-energy tensor.

The key features of 3-volumes and the stress-energy tensor are encapsulated by the above three-step procedure. But encapsulation is not sufficient; deep understanding is also required. To gain it, one must study special cases, both of 3-volumes and of the operation of the stress-energy machinery.

A Special Case

Interior of a soap box:

A soap box moves through spacetime. A man at an event $\mathscr{P}_0$ on the box's world line peers inside it, and examines all the soap, air, and electromagnetic fields it contains. He adds up all their 4-momenta to get a grand total $\boldsymbol{p}_{\text{box at }\mathscr{P}_0}$. How much is this grand total? One can calculate it by noting that the 4-momentum inside the box at $\mathscr{P}_0$ is precisely the 4-momentum crossing the box from past toward future there (Figure 5.1). Hence, the 4-momentum the man measures is

$$\boldsymbol{p}_{\text{box at }\mathscr{P}_0} = \boldsymbol{T}(\ldots, \boldsymbol{\Sigma}), \tag{5.4}$$

Its volume 1-form

where $\boldsymbol{\Sigma}$ is the box's volume 1-form at $\mathscr{P}_0$. But for such a soap box, $\boldsymbol{\Sigma}$ has a magnitude equal to the box's volume V as measured by a man in its momentary rest frame, and the box itself lies in one of the hyperplanes of $\boldsymbol{\Sigma}$; equivalently,

$$\boldsymbol{\Sigma} = -V\boldsymbol{u}, \tag{5.5}$$

where $\boldsymbol{u}$ is the soap box's 4-velocity at $\mathscr{P}_0$ (minus sign because $\boldsymbol{u}$, regarded as a 1-form, has positive sense toward the past, $u_0 < 0$); see Box 5.2. Hence, the total 4-momentum inside the box is

Its 4-momentum content

$$\boldsymbol{p}_{\text{box at }\mathscr{P}_0} = \boldsymbol{T}(\ldots, -V\boldsymbol{u}) = -V\boldsymbol{T}(\ldots, \boldsymbol{u}), \tag{5.6}$$

or, in component notation,

$$(p^\alpha)_{\text{box at }\mathscr{P}_0} = -VT^{\alpha\beta}u_\beta. \tag{5.6$'$}$$

The energy in the box, as measured in its rest frame, is minus the projection of the 4-momentum on the box's 4-velocity:

$$E = -\boldsymbol{u}\cdot\boldsymbol{p}_{\text{box at }\mathscr{P}_0} = +VT^{\alpha\beta}u_\alpha u_\beta = V\boldsymbol{T}(\boldsymbol{u}, \boldsymbol{u});$$

so

Its energy density

$$\begin{pmatrix}\text{energy density as}\\ \text{measured in box's}\\ \text{rest frame}\end{pmatrix} = \frac{E}{V} = \boldsymbol{T}(\boldsymbol{u}, \boldsymbol{u}). \tag{5.7}$$

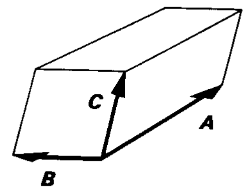
Another Special Case

A man riding with the same soap box opens its top and pours out some soap. In a very small interval of time $\Delta\tau$, how much total 4-momentum flows out of the box?

Box 5.2 THREE-DIMENSIONAL VOLUMES

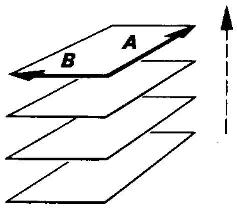
A. General Parallelepiped

1. *Edges* of parallelepiped are three vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$. One must order the edges; e.g., "$\boldsymbol{A}$ is followed by $\boldsymbol{B}$ is followed by $\boldsymbol{C}$."



(One dimension, that orthogonal to the parallelepiped, is suppressed here.)

2. *Volume trivector* is defined to be $\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C}$. It enters into the sophisticated theory of volumes (Chapter 4), but is not used much in the elementary theory.
3. *Volume 1-form* is defined by $\Sigma_\mu = \epsilon_{\mu\alpha\beta\gamma}A^\alpha B^\beta C^\gamma$. ($\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ must appear here in standard order as chosen in step 1.) Note that the vector "corresponding" to $\boldsymbol{\Sigma}$ and the volume trivector are related by $\boldsymbol{\Sigma} = -{}^*(\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C})$.
4. *Orientation* of the volume is defined to agree with the orientation of its 1-form $\boldsymbol{\Sigma}$. More specifically: the edges $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ lie in a hyperplane of $\boldsymbol{\Sigma}(\langle \boldsymbol{\Sigma}, \boldsymbol{A}\rangle = \langle \boldsymbol{\Sigma}, \boldsymbol{B}\rangle = \langle \boldsymbol{\Sigma}, \boldsymbol{C}\rangle = 0$; no "bongs of bell"). Thus, *the volume itself is one of $\boldsymbol{\Sigma}$'s hyperplanes!* The positive sense moving away from the volume is defined to be the positive sense of $\boldsymbol{\Sigma}$. *Note:* reversing the order of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ reverses the positive sense!



(One dimension, that along which $\boldsymbol{C}$ extends, is suppressed here.)

5. *The "standard orientation"* for a spacelike 3-volume has the positive sense of the 1-form $\boldsymbol{\Sigma}$ toward the future, corresponding to $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ forming a righthanded triad of vectors.

B. 3-Volumes of Arbitrary Shape

Can be analyzed by being broken up into union of parallelepipeds.

C. Interior of a Soap Box (Example)

1. *Analysis in soap box's rest frame.* Pick an event on the box's world line. The box's three edges there are three specific vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. In the box's rest frame they are purely spatial: $A^0 = B^0 = C^0 = 0$. Hence, the volume 1-form has components $\Sigma_j = 0$ and

Box 5.2 (*continued*)

$$\Sigma_0 = \epsilon_{0ijk}A^iB^jC^k = \det \begin{Vmatrix} A^1 & A^2 & A^3 \\ B^1 & B^2 & B^3 \\ C^1 & C^2 & C^3 \end{Vmatrix}$$

$= \boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C})$, in the standard notation of 3-dimensional vector analysis;

$= +V$ ($V =$ volume of box) if $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ are righthand ordered (positive sense of $\boldsymbol{\Sigma}$ toward future; standard orientation);

$= -V$ ($V =$ volume of box) if $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ are lefthand ordered (positive sense of $\boldsymbol{\Sigma}$ toward past).

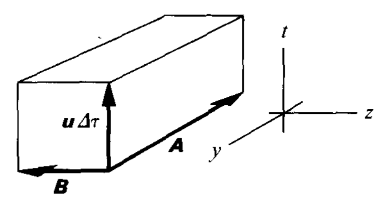
2. This result reexpressed in geometric language: Let $\boldsymbol{u}$ be the box's 4-velocity and V be its volume, as measured in its rest frame. Then either

$\boldsymbol{\Sigma} = -V\boldsymbol{u}$, in which case the "positive side" of the box's 3-surface is the future side, and its edges are ordered in a righthanded manner—the standard orientation;

or else

$\boldsymbol{\Sigma} = +V\boldsymbol{u}$, in which case the "positive side" is the past side, and the box's edges are ordered in a lefthanded manner.

D. 3-Volume Swept Out in Time $\Delta\tau$ by Two-Dimensional Top of a Soap Box (Example)



1. *Analysis in box's rest frame:* Pick an event on box's world line. There the two edges of the box top are vectors $\boldsymbol{A}$ and $\boldsymbol{B}$. In the box's rest frame, orient the space axes so that $\boldsymbol{A}$ and $\boldsymbol{B}$ lie in the y,z-plane. During the lapse of a proper time $\Delta\tau$, the box top sweeps out a 3-volume whose third edge is $\boldsymbol{u}\,\Delta\tau$ ($\boldsymbol{u} =$ 4-velocity of box). In the box's rest-frame, with ordering "$\boldsymbol{A}$ followed by $\boldsymbol{B}$ followed by $\boldsymbol{u}\,\Delta\tau$," the volume 1-form has components

$\Sigma_0 = \Sigma_2 = \Sigma_3 = 0$, and

$\Sigma_1 = \epsilon_{1jk0}A^jB^k\,\Delta\tau u^0 = -\epsilon_{01jk}A^jB^k\,\Delta\tau$

$= -\mathcal{A}\,\Delta\tau$ ($\mathcal{A} =$ area of box top) if $(\boldsymbol{e}_x, \boldsymbol{A}, \boldsymbol{B})$ are righthand ordered

$= +\mathcal{A}\,\Delta\tau$ ($\mathcal{A} =$ area of box top) if $(\boldsymbol{e}_x, \boldsymbol{A}, \boldsymbol{B})$ are lefthand ordered.

(Note: No standard orientation can be defined in this case, because $\boldsymbol{\Sigma}$ can be carried continuously into $-\boldsymbol{\Sigma}$ by purely spatial rotations.)

2. *This result reexpressed in geometric language:* Let $\mathcal{A}$ be the area of the box top as measured in its rest frame; and let $\boldsymbol{\sigma}$ be a unit 1-form, one of whose surfaces contains the box top and its 4-velocity (i.e., contains the box top's "world sheet"). Orient the positive sense of $\boldsymbol{\sigma}$ with the (arbitrarily chosen) positive sense of the box-top 3-volume. Then

$$\boldsymbol{\Sigma} = \mathcal{A}\,\Delta\tau\,\boldsymbol{\sigma}.$$

To answer this question, consider the three-dimensional volume swept out during $\Delta\tau$ by the box's opened two-dimensional top ("world sheet of top"). The 4-momentum asked for is the 4-momentum that crosses this world sheet in the positive sense (see Figure 5.1); hence, it is

The top of a soap box:

$$\boldsymbol{p}_{\text{flows out}} = \boldsymbol{T}(\ldots, \boldsymbol{\Sigma}), \tag{5.8}$$

where $\boldsymbol{\Sigma}$ is the world sheet's volume 1-form. Let $\mathcal{A}$ be the area of the box top, and $\boldsymbol{\sigma}$ be the outward-oriented unit 1-form, whose surfaces contain the world sheet (i.e., contain the box top and its momentary 4-velocity vector). Then

Its volume 1-form

$$\boldsymbol{\Sigma} = \mathcal{A}\, \Delta\tau\, \boldsymbol{\sigma} \tag{5.9}$$

(see Box 5.2); so the 4-momentum that flows out during $\Delta\tau$ is

Its 4-momentum that flows across

$$\boldsymbol{p}_{\text{flows out}} = \mathcal{A}\, \Delta\tau\, \boldsymbol{T}(\ldots, \boldsymbol{\sigma}). \tag{5.10}$$

§5.3. COMPONENTS OF STRESS-ENERGY TENSOR

Like all other tensors, the stress-energy tensor is a machine whose definition and significance transcend coordinate systems and reference frames. But any one observer, locked as he is into some one Lorentz frame, pays more attention to the components of $\boldsymbol{T}$ than to $\boldsymbol{T}$ itself. To each component he ascribes a specific physical significance. Of greatest interest, perhaps, is the "time-time" component. It is the total density of mass-energy as measured in the observer's Lorentz frame:

Physical interpretation of stress-energy tensor's components:

$$T_{00} = -T_0{}^0 = T^{00} = \boldsymbol{T}(\boldsymbol{e}_0, \boldsymbol{e}_0) = \text{density of mass-energy} \tag{5.11}$$

T^{00}: energy density

(cf. equation 5.7, with the observer's 4-velocity $\boldsymbol{u}$ replaced by the basis vector $\boldsymbol{e}_0 = \boldsymbol{u}$).

The "spacetime" components T^{j0} can be interpreted by considering the interior of a soap box at rest in the observer's frame. If its volume is V, then its volume 1-form is $\boldsymbol{\Sigma} = -V\boldsymbol{u} = +V\,\boldsymbol{d}t$; and the μ-component of 4-momentum inside it is

$$p^\mu = \langle \boldsymbol{d}x^\mu, \boldsymbol{p}\rangle = \boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{\Sigma}) = V\boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{d}t) = VT^{\mu 0}.$$

Thus, the 4-momentum per unit volume is

$$p^\mu/V = T^{\mu 0}, \tag{5.12a}$$

or, equivalently:

$$\begin{aligned} T^{00} = {} & \text{density of mass-energy} \\ & (\text{units: g/cm}^3\text{, or erg/cm}^3\text{, or cm}^{-2}); \end{aligned} \tag{5.13a}$$

$$\begin{aligned} T^{j0} = {} & \text{density of j-component of momentum} \\ & (\text{units: g (cm/sec) cm}^{-3}\text{, or cm}^{-2}). \end{aligned} \tag{5.13b}$$

T^{j0}: momentum density

The components $T^{\mu k}$ can be interpreted using a two-dimensional surface of area $\mathcal{A}$, at rest in the observer's frame with positive normal pointing in the k-direction.

During the lapse of time Δt, this 2-surface sweeps out a 3-volume with volume 1-form $\boldsymbol{\Sigma} = \mathcal{A}\,\Delta t\,\boldsymbol{d}x^k$ (see Box 5.2). The μ-component of 4-momentum that crosses the 2-surface in time Δt is

$$p^\mu = \boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{\Sigma}) = \mathcal{A}\,\Delta t\,\boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{d}x^k) = \mathcal{A}\,\Delta t\,T^{\mu k}.$$

Thus, the flux of 4-momentum (4-momentum crossing a unit surface oriented perpendicular to $\boldsymbol{e}_k$, in unit time) is

$$(p^\mu/\mathcal{A}\,\Delta t)_{\text{crossing surface} \perp \text{to } \boldsymbol{e}_k} = T^{\mu k}, \tag{5.12b}$$

or, equivalently:

T^{0k}: energy flux

T^{jk}: stress

T^{0k} = k-component of energy flux (units: erg/cm^2 sec, or cm^{-2}); (5.13c)

T^{jk} = j, k component of "stress"
$\equiv$ k-component of flux of j-component of momentum (5.13d)
$\equiv$ j-component of force produced by fields and matter at $x^k - \epsilon$ acting on fields and matter at $x^k + \epsilon$ across a unit surface, the perpendicular to which is $\boldsymbol{e}_k$
(units: dynes/cm^2, or cm^{-2}).

(Recall that "momentum transfer per second" is the same as "force.")

The stress-energy tensor is necessarily symmetric, $T^{\alpha\beta} = T^{\beta\alpha}$; but the proof of this will be delayed until several illustrations have been examined.

§5.4. STRESS-ENERGY TENSOR FOR A SWARM OF PARTICLES

Consider a swarm of particles. Choose some event $\mathcal{P}$ inside the swarm. Divide the particles near $\mathcal{P}$ into categories, $A = 1, 2, \ldots$, in such a way that all particles in the same category have the same properties:

$m_{(A)}$,	rest mass;
$\boldsymbol{u}_{(A)}$,	4-velocity;
$\boldsymbol{p}_{(A)} = m_{(A)}\boldsymbol{u}_{(A)}$,	4-momentum.

Number-flux vector for swarm of particles defined

Let $N_{(A)}$ be the number of category-A particles per unit volume, as measured in the particles' own rest frame. Then the *"number-flux vector"* $\boldsymbol{S}_{(A)}$, defined by

$$\boldsymbol{S}_{(A)} \equiv N_{(A)}\boldsymbol{u}_{(A)}, \tag{5.14}$$

has components with simple physical meanings. In a frame where category-A particles have ordinary velocity $\boldsymbol{v}_{(A)}$, the meanings are:

$$S^0_{(A)} = N_{(A)} u^0_{(A)} = \underbrace{N_{(A)}}_{\substack{\text{Number density in} \\ \text{particles' rest frame}}} \underbrace{[1 - v_{(A)}{}^2]^{-1/2}}_{\substack{\text{Lorentz contraction} \\ \text{factor for volume}}} = \text{number density}; \tag{5.15a}$$

$$\boldsymbol{S}_{(A)} = N_{(A)} \boldsymbol{u}_{(A)} = S^0_{(A)} \boldsymbol{v}_{(A)} = \text{flux of particles}. \tag{5.15b}$$

Consequently, the 4-momentum density has components

$$\begin{aligned} T^{\mu 0}_{(A)} &= p^\mu_{(A)} S^0_{(A)} = m_{(A)} u^\mu_{(A)} N_{(A)} u^0_{(A)} \\ &= m_{(A)} N_{(A)} u^\mu_{(A)} u^0_{(A)}; \end{aligned}$$

and the flux of μ-component of momentum across a surface with perpendicular direction $\boldsymbol{e}_j$ is

$$\begin{aligned} T^{\mu j}_{(A)} &= p^\mu_{(A)} S^j_{(A)} = m_{(A)} u^\mu_{(A)} N_{(A)} u^j_{(A)} \\ &= m_{(A)} N_{(A)} u^\mu_{(A)} u^j_{(A)}. \end{aligned}$$

These equations are precisely the $\mu, 0$ and μ, j components of the geometric, frame-independent equation

Stress-energy tensor for swarm of particles

$$\boldsymbol{T}_{(A)} = m_{(A)} N_{(A)} \boldsymbol{u}_{(A)} \otimes \boldsymbol{u}_{(A)} = \boldsymbol{p}_{(A)} \otimes \boldsymbol{S}_{(A)}. \tag{5.16}$$

The total number-flux vector and stress-energy tensor for all particles in the swarm near $\mathscr{P}$ are obtained by summing over all categories:

$$\boldsymbol{S} = \sum_A N_{(A)} \boldsymbol{u}_{(A)}; \tag{5.17}$$

$$\boldsymbol{T} = \sum_A m_{(A)} N_{(A)} \boldsymbol{u}_{(A)} \otimes \boldsymbol{u}_{(A)} = \sum_A \boldsymbol{p}_{(A)} \otimes \boldsymbol{S}_{(A)}. \tag{5.18}$$

§5.5. STRESS-ENERGY TENSOR FOR A PERFECT FLUID

Ideal gas defined

There is no simpler example of a fluid than a gas of noninteracting particles ("ideal gas") in which the velocities of the particles are distributed isotropically. In the Lorentz frame where isotropy obtains, symmetry argues equality of the diagonal space-space components of the stress-energy tensor,

$$T_{xx} = T_{yy} = T_{zz} = \sum_A \frac{m_{(A)} v_{x(A)}}{(1 - v_{(A)}{}^2)^{1/2}} \frac{N_{(A)} v_{x(A)}}{(1 - v_{(A)}{}^2)^{1/2}}, \tag{5.19}$$

and vanishing of all the off-diagonal components. Moreover, (5.19) represents a product: the number of particles per unit volume, multiplied by velocity in the x-direction (giving flux in the x-direction) and by momentum in the x-direction, giving the standard kinetic-theory expression for the pressure, p. Therefore, the stress-energy tensor takes the form

$$T_{\alpha\beta} = \begin{Vmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{Vmatrix} \tag{5.20}$$

in this special Lorentz frame—the "rest frame" of the gas. Here the quantity ρ has nothing directly to do with the rest-masses of the constituent particles. It measures the density of rest-plus-kinetic energy of these particles.

Rewrite (5.20) in terms of the 4-velocity $u^\alpha = (1, 0, 0, 0)$ of the fluid in the gas's rest frame, and find

$$\begin{aligned} T_{\alpha\beta} &= \begin{Vmatrix} \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix} + \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{Vmatrix} \\ &= \rho u_\alpha u_\beta + p(\eta_{\alpha\beta} + u_\alpha u_\beta), \end{aligned}$$

or, in frame-independent, geometric language

Stress-energy tensor for ideal gas or perfect fluid

$$T = p\boldsymbol{g} + (\rho + p)\boldsymbol{u} \otimes \boldsymbol{u}. \tag{5.21}$$

Perfect fluid defined

Expression (5.21) has general application. It is exact for the "ideal gas" just considered. It is also exact for any fluid that is "perfect" in the sense that it is free of such transport processes as heat conduction and viscosity, and therefore (in the rest frame) free of shear stress (diagonal stress tensor; diagonal components identical, because if they were not identical, a rotation of the frame of reference would reveal presence of shear stress). However, for a general perfect fluid, density ρ of mass-energy as measured in the fluid's rest frame includes not only rest mass plus kinetic energy of particles, but also energy of compression, energy of nuclear binding, and all other sources of mass-energy [total density of mass-energy as it might be determined by an idealized experiment, such as that depicted in Figure 1.12, with the sample mass at the center of the sphere, and the test particle executing oscillations of small amplitude about that location, with $\omega^2 = (4\pi/3)\rho$].

§5.6. ELECTROMAGNETIC STRESS-ENERGY

Faraday, with his picture of tensions along lines of force and pressures at right angles to them (Figure 5.2), won insight into new features of electromagnetism. In addition to the tension $\boldsymbol{E}^2/8\pi$ (or $\boldsymbol{B}^2/8\pi$) along lines of force, and an equal pressure at right angles, one has the Poynting flux $(\boldsymbol{E} \times \boldsymbol{B})/4\pi$ and the Maxwell expression for the

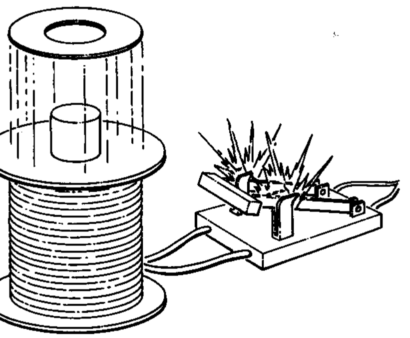


Figure 5.2.
Faraday stresses at work. When the electromagnet is connected to an alternating current, the aluminum ring flies into the air.

energy density, $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$. All these quantities find their places in the Maxwell stress-energy tensor, defined by

$$4\pi T^{\mu\nu} = F^{\mu\alpha}F^{\nu}{}_{\alpha} - \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}. \tag{5.22}$$

Stress-energy tensor for electromagnetic field

Exercise 5.1.

EXERCISE

Show that expression (5.22), evaluated in a Lorentz coordinate frame, gives

$$\begin{aligned} T^{00} &= (\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi, \qquad T^{0j} = T^{j0} = (\boldsymbol{E} \times \boldsymbol{B})^j/4\pi, \\ T^{jk} &= \frac{1}{4\pi}\left[-(E^jE^k + B^jB^k) + \frac{1}{2}(\boldsymbol{E}^2 + \boldsymbol{B}^2)\,\delta^{jk}\right]. \end{aligned} \tag{5.23}$$

Show that the stress tensor does describe a tension $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$ along the field lines and a pressure $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$ perpendicular to the field lines, as stated in the text.

---

§5.7. SYMMETRY OF THE STRESS-ENERGY TENSOR

All the stress-energy tensors explored above were symmetric. That they could not have been otherwise one sees as follows.

Proof that stress-energy tensor is symmetric

Calculate in a specific Lorentz frame. Consider first the momentum density (components T^{j0}) and the energy flux (components T^{0j}). They must be equal because energy = mass ("$E = Mc^2 = M$"):

$$\begin{aligned} T^{0j} &= (\text{energy flux}) \\ &= (\text{energy density}) \times (\text{mean velocity of energy flow})^j \\ &= (\text{mass density}) \times (\text{mean velocity of mass flow})^j \\ &= (\text{momentum density}) = T^{j0}. \end{aligned} \tag{5.24}$$

Only the stress tensor T^{jk} remains. For it, one uses the same standard argument as in Newtonian theory. Consider a very small cube, of side L, mass-energy $T^{00}L^3$,

2

and moment of inertia $\sim T^{00}L^5$. With the space coordinates centered at the cube, the expression for the z-component of torque exerted on the cube by its surroundings is

$$
\begin{aligned}
\tau^z = \; & \underbrace{(-T^{yx}L^2)}_{\begin{pmatrix}y\text{-component}\\ \text{of force on}\\ +x \text{ face}\end{pmatrix}} \underbrace{(L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ +x \text{ face}\end{pmatrix}} + \underbrace{(T^{yx}L^2)}_{\begin{pmatrix}y\text{-component}\\ \text{of force on}\\ -x \text{ face}\end{pmatrix}} \underbrace{(-L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ -x \text{ face}\end{pmatrix}} \\
& - \underbrace{(-T^{xy}L^2)}_{\begin{pmatrix}x\text{-component}\\ \text{of force on}\\ +y \text{ face}\end{pmatrix}} \underbrace{(L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ +y \text{ face}\end{pmatrix}} - \underbrace{(T^{xy}L^2)}_{\begin{pmatrix}x\text{-component}\\ \text{of force on}\\ -y \text{ face}\end{pmatrix}} \underbrace{(-L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ -y \text{ face}\end{pmatrix}} \\
= \; & (T^{xy} - T^{yx})L^3.
\end{aligned}
$$

Since the torque decreases only as L^3 with decreasing L, while the moment of inertia decreases as L^5, the torque will set an arbitrarily small cube into arbitrarily great angular acceleration—which is absurd. To avoid this, the stresses distribute themselves so the torque vanishes:

$$T^{yx} = T^{xy}.$$

Put differently, if the stresses were not so distributed, the resultant infinite angular accelerations would instantaneously redistribute them back to equilibrium. This condition of torque balance, repeated for all other pairs of directions, is equivalent to symmetry of the stresses:

$$T^{jk} = T^{kj}. \tag{5.25}$$

§5.8. CONSERVATION OF 4-MOMENTUM: INTEGRAL FORMULATION

Energy-momentum conservation has been a cornerstone of physics for more than a century. Nowhere does its essence shine forth so clearly as in Einstein's geometric formulation of it (Figure 5.3,a). There one examines a four-dimensional region of spacetime $\mathcal{V}$ bounded by a closed, three-dimensional surface $\partial\mathcal{V}$. As particles and fields flow into $\mathcal{V}$ and later out, they carry 4-momentum. Inside $\mathcal{V}$ the particles collide, break up, radiate; radiation propagates, jiggles particles, produces pairs. But at each stage in this complex maze of physical processes, total energy-momentum is conserved. The energy-momentum lost by particles goes into fields; the energy-momentum lost by fields goes into particles. So finally, when the "river" of 4-momentum exits from $\mathcal{V}$, it carries out precisely the same energy-momentum as it carried in.

Integral conservation law for 4-momentum:
$\oint_{\partial\mathcal{V}} \boldsymbol{T}\cdot d^3\boldsymbol{\Sigma} = 0$

Restate this equality by asking for the total flux of 4-momentum *outward* across $\partial\mathcal{V}$. Count inflowing 4-momentum negatively. Then "inflow equals outflow" means "total outflow vanishes":

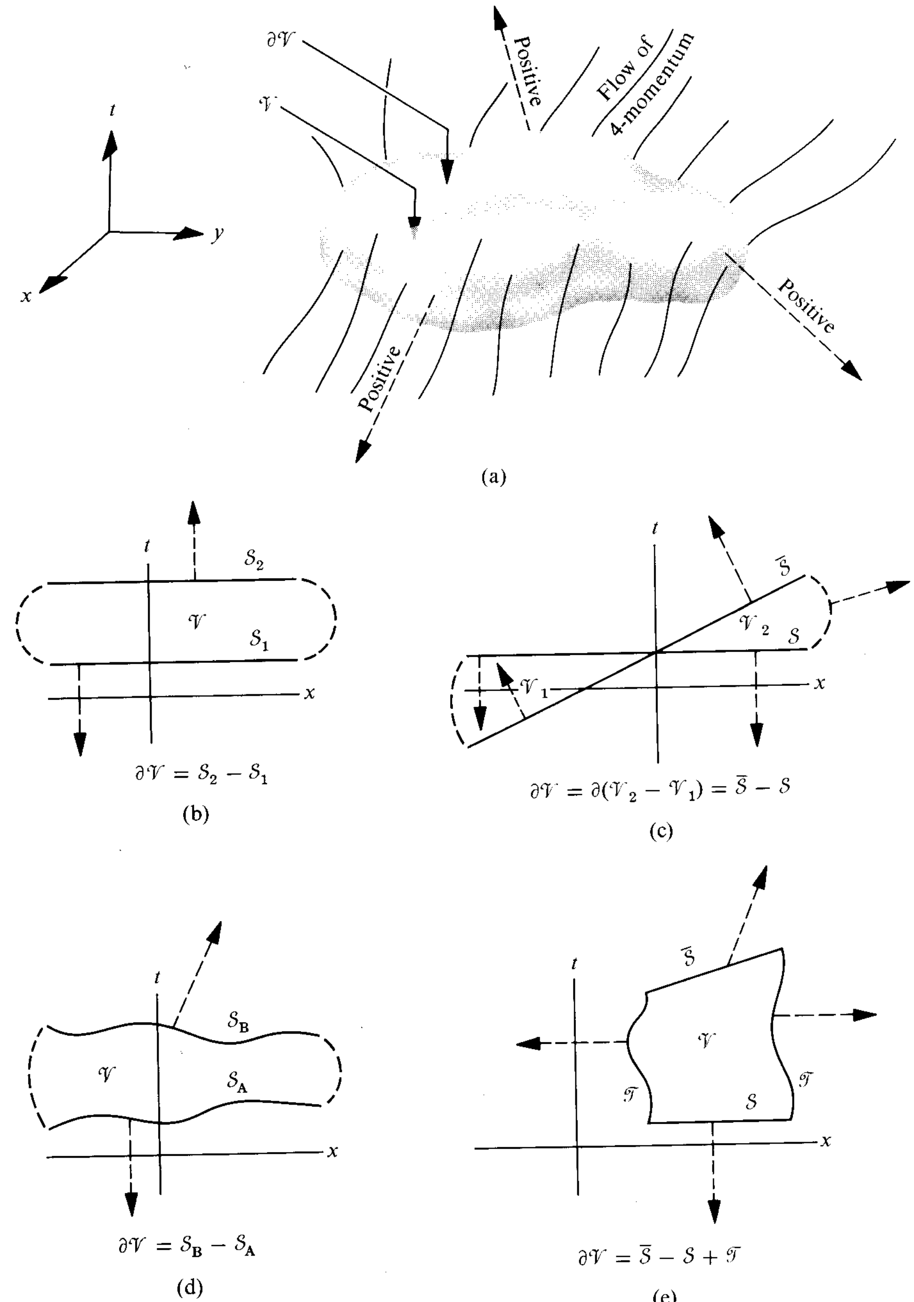


Figure 5.3.
(a) A four-dimensional region of spacetime $\mathcal{V}$ bounded by a closed three-dimensional surface $\partial\mathcal{V}$. The positive sense of $\partial\mathcal{V}$ is defined to be everywhere outward (away from $\mathcal{V}$). Conservation of energy-momentum demands that every bit of 4-momentum which flows into $\mathcal{V}$ through $\partial\mathcal{V}$ must somewhere flow back out; none can get lost inside; the interior contains no "sinks." Equivalently, the total flux of 4-momentum across $\partial\mathcal{V}$ in the positive (outward) sense must be zero:

$$\oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_\alpha = 0.$$

Figures (b), (c), (d), and (e) depict examples to which the text applies this law of conservation of 4-momentum. All symbols $\mathcal{V}$ (or S) in these figures mean spacetime volumes (or spacelike 3-volumes) with standard orientations. The dotted arrows indicate the positive sense of the closed surface $\partial\mathcal{V}$ used in the text's discussion of 4-momentum conservation. How $\partial\mathcal{V}$ is constructed from the surfaces S and $\mathcal{T}$ is indicated by formulas below the figures. For example, in case (b), $\partial\mathcal{V} = S_2 - S_1$ means that $\partial\mathcal{V}$ is made by joining together S_2 with its standard orientation and S_1 with reversed orientation.

Total flux of 4-momentum outward across a closed three-dimensional surface must vanish. (5.26)

To calculate the total outward flux in the most elementary of fashions, approximate the closed 3-surface $\partial\mathcal{V}$ by a large number of flat 3-volumes ("boiler plates") with positive direction oriented outward (away from $\mathcal{V}$). Then

$$\boldsymbol{p}_{\text{total out}} = \sum_{\text{boiler plates } A} \boldsymbol{T}(\ldots, \boldsymbol{\Sigma}_{(A)}) = 0, \tag{5.27}$$

where $\boldsymbol{\Sigma}_{(A)}$ is the volume 1-form of boiler plate A. Equivalently, in component notation

$$p^{\mu}{}_{\text{total out}} = \sum_{A} T^{\mu\alpha} \Sigma_{(A)\alpha}. \tag{5.27$'$}$$

To be slightly more sophisticated about the calculation, take the limit as the number of boiler plates goes to infinity and their sizes go to zero. The result is an integral (Box 5.3, at the end of this section),

$$p^{\mu}{}_{\text{total out}} = \oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_{\alpha} = 0. \tag{5.28}$$

Think of this (like all component equations) as a convenient way to express a coordinate-independent statement:

$$\boldsymbol{p}_{\text{total out}} = \oint_{\partial\mathcal{V}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma} = 0. \tag{5.29}$$

To be more sophisticated yet (not recommended on first reading of this book) and to simplify the computations in practical cases, interpret the integrands as exterior differential forms (Box 5.4, at the end of this section).

But however one calculates it, and however one interprets the integrands, the statement of the result is simple: the total flux of 4-momentum outward across a closed 3-surface must vanish.

Special cases of integral conservation law:

Several special cases of this "integral conservation law," shown in Figure 5.3, are instructive. There shown, in addition to the general case (a), are:

Case (b)

The closed 3-surface $\partial\mathcal{V}$ is made up of two slices taken at constant time t of a specific Lorentz frame, plus timelike surfaces at "infinity" that join the two slices together. The surfaces at infinity do not contribute to $\oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_{\alpha}$ if the stress-energy tensor dies out rapidly enough there. The boundary $\partial\mathcal{V}$ of the standard-oriented 4-volume $\mathcal{V}$, by definition, has its positive sense away from $\mathcal{V}$. This demands nonstandard orientation of $\mathcal{S}_1$ (positive sense toward past), as is indicated by writing $\partial\mathcal{V} = \mathcal{S}_2 - \mathcal{S}_1$; and it produces a sign flip in the evaluation of the hypersurface integral

$$0 = \oint_{\partial\mathcal{V}} T^{\alpha\mu}\, d^3\Sigma_\mu = -\int_{\mathcal{S}_1} T^{\alpha 0}\, dx\, dy\, dz + \int_{\mathcal{S}_2} T^{\alpha 0}\, dx\, dy\, dz.$$

Because $T^{\alpha 0}$ is the density of 4-momentum, this equation says

Total 4-momentum conserved in time

$$\begin{aligned}\begin{pmatrix}\text{total 4-momentum in}\\ \text{all of space at time } t_1\end{pmatrix} &= \int_{\mathcal{S}_1} T^{\alpha 0}\, dx\, dy\, dz \\ &= \begin{pmatrix}\text{total 4-momentum in}\\ \text{all of space at time } t_2\end{pmatrix} = \int_{\mathcal{S}_2} T^{\alpha 0}\, dx\, dy\, dz.\end{aligned} \tag{5.30}$$

Case (c)

Here one wants to compare hypersurface integrals over $\mathcal{S}$ and $\overline{\mathcal{S}}$, which are slices of constant time, $t = \text{const}$ and $\bar{t} = \text{const}$ in two different Lorentz frames. To form a closed surface, one adds time-like hypersurfaces at infinity and assumes they do not contribute to the integral. The orientations fit together smoothly and give a closed surface

$$\partial\mathcal{V} = \overline{\mathcal{S}} - \mathcal{S} + (\text{surfaces at infinity})$$

only if one takes $\mathcal{V} = \mathcal{V}_2 - \mathcal{V}_1$—i.e., only if one uses the nonstandard 4-volume orientation in $\mathcal{V}_1$. (See part A.1 of Box 5.3 for "standard" versus "non-standard" orientation.) The integral conservation law then gives

$$0 = \int_{\overline{\mathcal{S}}} \boldsymbol{T}\cdot d^3\boldsymbol{\Sigma} - \int_{\mathcal{S}} \boldsymbol{T}\cdot d^3\boldsymbol{\Sigma},$$

or, equivalently,

Total 4-momentum the same in all Lorentz frames

$$\begin{aligned}\int_{\overline{\mathcal{S}}} \boldsymbol{T}\cdot d^3\boldsymbol{\Sigma} &= (\text{total 4-momentum } \boldsymbol{p} \text{ on } \overline{\mathcal{S}}) \\ = \int_{\mathcal{S}} \boldsymbol{T}\cdot d^3\boldsymbol{\Sigma} &= (\text{total 4-momentum } \boldsymbol{p} \text{ on } \mathcal{S}).\end{aligned} \tag{5.31}$$

This says that observers in different Lorentz frames measure the same total 4-momentum $\boldsymbol{p}$. It does *not* mean that they measure the same components ($p^\alpha \neq p^{\bar{\alpha}}$); rather, it means they measure the same geometric vector

$$\boldsymbol{p}_{\text{on } \mathcal{S}} = p^\alpha \boldsymbol{e}_\alpha = \boldsymbol{p}_{\text{on } \overline{\mathcal{S}}} = p^{\bar{\alpha}} \boldsymbol{e}_{\bar{\alpha}},$$

a vector whose components are connected by the usual Lorentz transformation law

$$p^\alpha = \Lambda^\alpha{}_{\bar{\beta}}\, p^{\bar{\beta}}. \tag{5.32}$$

Case (d)

Here the contribution to the integral comes entirely from two arbitrary spacelike hypersurfaces, $\mathcal{S}_A$ and $\mathcal{S}_B$, cutting all the way across spacetime. As in cases (a) and (b), the integral form of the conservation law says

Total 4-momentum independent of hypersurface where measured

$$\boldsymbol{p}_{\text{on } \mathcal{S}_A} = \boldsymbol{p}_{\text{on } \mathcal{S}_B}; \tag{5.33}$$

i.e., *the total 4-momentum on a spacelike slice through spacetime is independent of the specific slice chosen*—so long as the energy-momentum flux across the "hypersurface at infinity" connecting $\mathcal{S}_A$ and $\mathcal{S}_B$ is zero.

Case (e)

This case concerns a box whose walls oscillate and accelerate as time passes. The three-dimensional boundary $\partial\mathcal{V}$ is made up of (1) the interior $\mathcal{S}$ of the box, at an initial moment of time $t =$ constant in the box's initial Lorentz frame, taken with nonstandard orientation; (2) the interior $\overline{\mathcal{S}}$ of the box, at $\bar{t} =$ constant in its final Lorentz frame, with standard orientation; (3) the 3-volume $\mathcal{T}$ swept out by the box's two-dimensional faces between the initial and final states, with positive sense oriented outward. The integral conservation law $\int_{\partial\mathcal{V}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma} = 0$ says

Change with time of 4-momentum in a box equals flux of 4-momentum across its faces

$$\begin{pmatrix}\text{total 4-momentum}\\ \text{in box at } \overline{\mathcal{S}}\end{pmatrix} - \begin{pmatrix}\text{total 4-momentum}\\ \text{in box at } \mathcal{S}\end{pmatrix} = \begin{pmatrix}\text{total 4-momentum that enters box through}\\ \text{its faces between states } \mathcal{S} \text{ and } \overline{\mathcal{S}}\end{pmatrix}. \tag{5.34}$$

§5.9. CONSERVATION OF 4-MOMENTUM: DIFFERENTIAL FORMULATION

Complementary to any "integral conservation law in flat spacetime" is a "differential conservation law" with identical information content. To pass back and forth between them, one can use Gauss's theorem.

Gauss's theorem in four dimensions, applied to the law of 4-momentum conservation, converts the surface integral of $T^{\mu\alpha}$ into a volume integral of $T^{\mu\alpha}{}_{,\alpha}$:

$$0 = \oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_\alpha = \int_{\mathcal{V}} T^{\mu\alpha}{}_{,\alpha}\, dt\, dx\, dy\, dz. \tag{5.35}$$

(See Box 5.3 for elementary discussion; Box 5.4 for sophisticated discussion.) If the integral of $T^{\mu\alpha}{}_{,\alpha}$ is to vanish, as demanded, for any and every 4-volume $\mathcal{V}$, then $T^{\mu\alpha}{}_{,\alpha}$ must itself vanish everywhere in spacetime:

Differential conservation law for 4-momentum: $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$

$$T^{\mu\alpha}{}_{,\alpha} = 0; \text{ i.e., } \boldsymbol{\nabla} \cdot \boldsymbol{T} = 0 \text{ everywhere.} \tag{5.36}$$

(continued on page 152)

Box 5.3 VOLUME INTEGRALS, SURFACE INTEGRALS, AND GAUSS'S THEOREM IN COMPONENT NOTATION

A. Volume Integrals in Spacetime

1. By analogy with three-dimensional space, the volume of a "hyperparallelepiped" with vector edges $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ is

$$\text{4-volume} \equiv \Omega \equiv \epsilon_{\alpha\beta\gamma\delta} A^\alpha B^\beta C^\gamma D^\delta = \det \begin{Vmatrix} A^0 & A^1 & A^2 & A^3 \\ B^0 & B^1 & B^2 & B^3 \\ C^0 & C^1 & C^2 & C^3 \\ D^0 & D^1 & D^2 & D^3 \end{Vmatrix}$$

$$= {}^*(\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C} \wedge \boldsymbol{D}).$$

Here, as for 3-volumes, orientation matters; interchange of any two edges reverses the sign of Ω. The *standard orientation* for any 4-volume is the one which makes Ω positive; thus, $\boldsymbol{e}_0 \wedge \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3$ has standard orientation if $\boldsymbol{e}_0$ points toward the future and $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ are a righthanded triad.

2. The "volume element" whose edges in a specific, standard-oriented Lorentz frame are

$$A^\alpha = (\Delta t, 0, 0, 0),\ B^\alpha = (0, \Delta x, 0, 0),\ C^\alpha = (0, 0, \Delta y, 0),\ D^\alpha = (0, 0, 0, \Delta z)$$

has a 4-volume, according to the above definition, given by

$$\Delta^4 \Omega = \epsilon_{0123}\, \Delta t\, \Delta x\, \Delta y\, \Delta z = \Delta t\, \Delta x\, \Delta y\, \Delta z.$$

3. Thus, the volume integral of a tensor $\boldsymbol{S}$ over a four-dimensional region $\mathcal{V}$ of spacetime, defined as

$$\boldsymbol{M} \equiv \lim_{\begin{pmatrix}\text{number of}\\ \text{elementary}\\ \text{volumes}\\ \to\infty\end{pmatrix}} \sum_{\begin{pmatrix}\text{elementary}\\ \text{volumes } \mathcal{U}\\ \text{in } \mathcal{V}\end{pmatrix}} \boldsymbol{S}_{\text{at center of } \mathcal{U}}\, (\text{volume of } \mathcal{U}),$$

can be calculated in a Lorentz frame by

$$M^\alpha{}_{\beta\gamma} = \int_{\mathcal{V}} S^\alpha{}_{\beta\gamma}\, d^4\Omega = \int_{\mathcal{V}} S^\alpha{}_{\beta\gamma}\, dt\, dx\, dy\, dz.$$

Box 5.3 ***(continued)***

B. Integrals over *3*-Surfaces in Spacetime

1. Introduce arbitrary coordinates a, b, c on the three-dimensional surface. The elementary volume bounded by coordinate surfaces

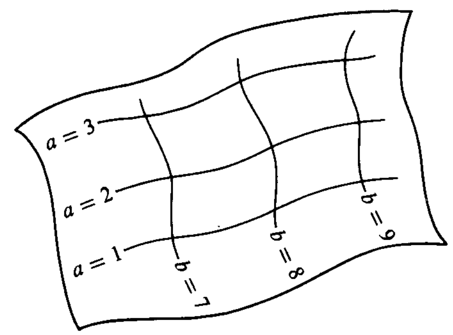
   $$a_0 < a < a_0 + \Delta a, \qquad b_0 < b < b_0 + \Delta b,$$
   $$c_0 < c < c_0 + \Delta c$$

   has edges

   $$A^\alpha = \frac{\partial x^\alpha}{\partial a}\Delta a,\ B^\beta = \frac{\partial x^\beta}{\partial b}\Delta b,\ C^\gamma = \frac{\partial x^\gamma}{\partial c}\Delta c;$$

   so its volume 1-form is

   $$\Delta^3\Sigma_\mu = \epsilon_{\mu\alpha\beta\gamma}\frac{\partial x^\alpha}{\partial a}\frac{\partial x^\beta}{\partial b}\frac{\partial x^\gamma}{\partial c}\Delta a\,\Delta b\,\Delta c.$$



2. The integral of a tensor $\boldsymbol{S}$ over the 3-surface $\mathcal{S}$ thus has components

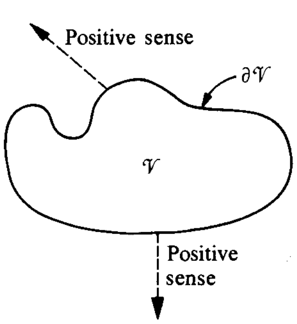
   $$N^\alpha{}_\beta = \int_{\mathcal{S}} S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma = \int_{\mathcal{S}} S^\alpha{}_\beta{}^\gamma \epsilon_{\gamma\mu\nu\lambda}\frac{\partial x^\mu}{\partial a}\frac{\partial x^\nu}{\partial b}\frac{\partial x^\lambda}{\partial c}\, da\, db\, dc.$$

   An equivalent formula involving a Jacobian is often used (see exercise 5.5):

   $$N^\alpha{}_\beta = \int_{\mathcal{S}} S^\alpha{}_\beta{}^\gamma \frac{1}{3!}\epsilon_{\gamma\mu\nu\lambda}\frac{\partial(x^\mu, x^\nu, x^\lambda)}{\partial(a, b, c)}\, da\, db\, dc.$$

C. Gauss's Theorem Stated

1. Consider a bounded four-dimensional region of spacetime $\mathcal{V}$ with closed boundary $\partial\mathcal{V}$. Orient the volume 1-forms on $\partial\mathcal{V}$ so that the "positive sense" is away from $\mathcal{V}$.
2. Choose a tensor field $\boldsymbol{S}$. Integrate its divergence over $\mathcal{V}$, and integrate it itself over $\partial\mathcal{V}$. The results must be the same (*Gauss's theorem*):



$$\int_{\mathcal{V}} S^\alpha{}_\beta{}^\gamma{}_{,\gamma}\, d^4\Omega = \oint_{\partial\mathcal{V}} S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma.$$

D. Proof of Gauss's Theorem

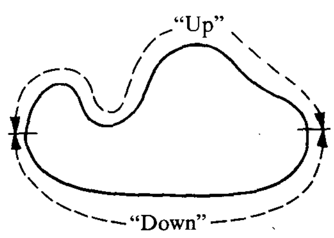
1. The indices α and β of $S^{\alpha}{}_{\beta}{}^{\gamma}$ "go along for a free ride," so one can suppress them from the proof. Then the equation to be derived is

$$\int_{\mathscr{V}} S^{\gamma}{}_{,\gamma}\, dt\, dx\, dy\, dz = \oint_{\partial\mathscr{V}} S^{\gamma}\, d^3\Sigma_{\gamma}.$$

2. Since the integral of a derivative is just the original function, the volume integral of $S^{0}{}_{,0}$ is

$$\int_{\mathscr{V}} S^{0}{}_{,0}\, dt\, dx\, dy\, dz$$

$$= \int_{\text{"up"}} S^0\, dx\, dy\, dz - \int_{\text{"down"}} S^0\, dx\, dy\, dz.$$



3. The surface integral $\int_{\partial\mathscr{V}} S^0\, d^3\Sigma_0$ can be reduced to the same set of terms:
   a. Use x, y, z as coordinates on $\partial\mathscr{V}$. On the "up" side, $d^3\Sigma_0$ must be positive to achieve a "positive" sense pointing away from $\mathscr{V}$, so (see part B above)

$$d^3\Sigma_0 = \epsilon_{0\alpha\beta\gamma}\frac{\partial x^\alpha}{\partial x}\frac{\partial x^\beta}{\partial y}\frac{\partial x^\gamma}{\partial z}\, dx\, dy\, dz = \epsilon_{0123}\, dx\, dy\, dz = dx\, dy\, dz.$$

   b. On the "down" side, $d^3\Sigma_0$ must be negative, so

$$d^3\Sigma_0 = -dx\, dy\, dz.$$

   c. Hence,

$$\int_{\partial\mathscr{V}} S^0\, d^3\Sigma_0 = \int_{\text{"up"}} S^0\, dx\, dy\, dz - \int_{\text{"down"}} S^0\, dx\, dy\, dz.$$

4. Equality is proved for the other components in the same manner. Adding components produces the result desired:

$$\int_{\mathscr{V}} S^{\gamma}{}_{,\gamma}\, d^4\Omega = \oint_{\partial\mathscr{V}} S^{\gamma}\, d^3\Sigma_{\gamma}.$$

2

FOR THE READER WHO HAS STUDIED CHAPTER 4

Box 5.4 I. EVERY INTEGRAL IS THE INTEGRAL OF A FORM.
II. THE THEOREM OF GAUSS IN THE LANGUAGE OF FORMS.

I. *Every integral encountered* in Chapter 5 can be interpreted as *the integral of an exterior differential form.* This circumstance shows up in fourfold and threefold integrals, for example, in the fact that

$$d^4\Omega = \boldsymbol{\varepsilon} = {}^*1 = \varepsilon_{0123}\, \boldsymbol{d}t \wedge \boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z$$

and

$$d^3\Sigma_\mu = \varepsilon_{\mu|\alpha\beta\gamma|}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma$$

are basis 4- and 3-forms. (Recall: the indices $\alpha\beta\gamma$ between vertical bars are to be summed only over $0 \leq \alpha < \beta < \gamma \leq 3$.) A more extensive glossary of notations is found in C below.

II. *Gauss's Theorem for a tensor integral* in flat space reads

$$\int_{\mathcal{V}} (\boldsymbol{\nabla} \cdot \boldsymbol{S})\, d^4\Omega = \oint_{\partial\mathcal{V}} \boldsymbol{S} \cdot d\boldsymbol{\Sigma}$$

for any tensor, such as $\boldsymbol{S} = S^\alpha{}_\beta{}^\gamma \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \otimes \boldsymbol{e}_\gamma$ (see Box 5.3 for component form). It is an application of the generalized Stokes Theorem (Box 4.1), and *depends on the fact that the basis vectors $\boldsymbol{e}_\alpha$ and $\boldsymbol{\omega}^\beta$ of a global Lorentz frame are constants,* i.e., are independent of x. The definitions follow in A; the proof is in B.

A. *Tensor-valued integrals* can be defined in flat spaces because one uses constant basis vectors. Thus one defines

$$\int \boldsymbol{S} \cdot d^3\boldsymbol{\Sigma} = \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$$

for a tensor of the indicated rank. One justifies pulling basis vectors and forms outside the integral sign because they are constants, independent of location in spacetime. Each of the numbers $\int S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$ (for $\alpha, \beta = 0, 1, 2, 3$) is then evaluated by substituting *any* properly oriented parametrization of the hypersurface into the 3-form $S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$ as described in Box 4.1 (arbitrary curvilinear parametrization in the part of the calculation not involving the "free indices" α and β). In other words, $\boldsymbol{S} \cdot d^3\boldsymbol{\Sigma} = \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \otimes S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$ is considered a "tensor-valued 3-form." Under an integral sign, it is contracted with the hyperplane element tangent to the 3-surface $\mathcal{P}(\lambda^1, \lambda^2, \lambda^3)$ of integration to form the integral

$$\int \boldsymbol{S} \cdot d^3\boldsymbol{\Sigma} = \int \left\langle \boldsymbol{S} \cdot d^3\boldsymbol{\Sigma}, \frac{\partial\mathcal{P}}{\partial\lambda^1} \wedge \frac{\partial\mathcal{P}}{\partial\lambda^2} \wedge \frac{\partial\mathcal{P}}{\partial\lambda^3} \right\rangle d\lambda^1\, d\lambda^2\, d\lambda^3$$

$$= \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int S^\alpha{}_\beta{}^\gamma \varepsilon_{\gamma|\lambda\mu\nu|} \underbrace{\frac{\partial(x^\lambda, x^\mu, x^\nu)}{\partial(\lambda^1, \lambda^2, \lambda^3)}}_{\substack{\text{Jacobian}\\ \text{determinant}}} d\lambda^1\, d\lambda^2\, d\lambda^3.$$

Although constant basis vectors $\boldsymbol{e}_\alpha$, $\boldsymbol{\omega}^\beta$ derived from rectangular coordinates are essential here, *a completely general parametrization of the hypersurface may be used.*

B. *The proof of Gauss's Theorem* is a computation:

$$\oint_{\partial\mathcal{V}} \boldsymbol{S}\cdot d^3\boldsymbol{\Sigma} = \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \oint_{\partial\mathcal{V}} S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma \qquad (\boldsymbol{e}_\alpha, \boldsymbol{\omega}^\beta \text{ are constant})$$

$$= \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int_{\mathcal{V}} \boldsymbol{d}(S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma) \qquad \text{(Stokes Theorem)}$$

$$= \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int_{\mathcal{V}} S^\alpha{}_\beta{}^\gamma{}_{,\gamma}\, {*1} \qquad \text{(see below)}$$

$$= \int_{\mathcal{V}} (\boldsymbol{\nabla}\cdot\boldsymbol{S})\, d^4\Omega. \qquad \text{(merely notation)}$$

The missing computational step above is

$$\begin{aligned} \boldsymbol{d}(S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma) &= (\partial S^\alpha{}_\beta{}^\gamma/\partial x^\rho)\, \boldsymbol{d}x^\rho \wedge d^3\Sigma_\gamma \\ &= (\partial S^\alpha{}_\beta{}^\gamma/\partial x^\gamma)\, {*1}. \end{aligned}$$

Here the first step uses $\boldsymbol{d}(d^3\Sigma_\gamma) = 0$ (which follows from $\varepsilon_{\mu\alpha\beta\gamma} = \text{const}$ in flat spacetime). The second step uses

$$\boldsymbol{d}x^\rho \wedge d^3\Sigma_\gamma = \delta^\rho_\gamma\, {*1}.$$

[Write the lefthand side of this identity as $\varepsilon_{\gamma|\mu\nu\lambda|}\, \boldsymbol{d}x^\rho \wedge \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\lambda$. The only possible non-zero term in the sum over $\mu\nu\lambda$ is the one with $\mu < \nu < \lambda$ all different from ρ. The righthand side is the value of this term.]

C. *Glossary of notations.*

Charge density 3-form:

$$\begin{aligned} {*\boldsymbol{J}} = J^\mu\, d^3\Sigma_\mu &= \boldsymbol{J}\cdot d^3\boldsymbol{\Sigma} \\ &= \underbrace{J^\mu \varepsilon_{\mu\alpha\beta\gamma}}_{(*J)_{\alpha\beta\gamma}} \overbrace{\phantom{\varepsilon_{\mu\alpha\beta\gamma}}\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma/3!}^{d^3\Sigma_\mu} \end{aligned}$$

Maxwell and Faraday 2-forms:

$${*\boldsymbol{F}} = \frac{1}{2} F^{\mu\nu}\, d^2S_{\mu\nu};$$

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu.$$

Basis 2-forms:

$$\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta; \qquad \text{(one way to label)}$$

$$d^2S_{\mu\nu} = \varepsilon_{\mu\nu|\alpha\beta|}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta. \qquad \text{(dual way to label)}$$

Energy-momentum density 3-form:

$$\boldsymbol{T}\cdot d^3\boldsymbol{\Sigma} \equiv \underbrace{\boldsymbol{e}_\mu T^{\mu\nu}\, d^3\Sigma_\nu \equiv {*\boldsymbol{T}};}_{\text{dual on last index, } (*T)^\mu{}_{\alpha\beta\gamma} = T^{\mu\nu}\varepsilon_{\nu\alpha\beta\gamma}.}$$

Angular momentum density 3-form:

$$\boldsymbol{\mathcal{J}}\cdot d^3\boldsymbol{\Sigma} \equiv \frac{1}{2}\boldsymbol{e}_\mu \wedge \boldsymbol{e}_\nu \mathcal{J}^{\mu\nu\alpha}\, d^3\Sigma_\alpha \equiv {*\boldsymbol{\mathcal{J}}};$$

$$(*\mathcal{J})^{\mu\nu}{}_{\alpha\beta\gamma} = \mathcal{J}^{\mu\nu\lambda}\varepsilon_{\lambda\alpha\beta\gamma}.$$

(In the frame-independent equation $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$, one need not worry about which slot of $\boldsymbol{T}$ to take the divergence on; the slots are symmetric, so either can be used.)

The equation $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ is *the differential formulation of the law of 4-momentum conservation*. It is also called the *equation of motion for stress-energy*, because it places constraints on the dynamic evolution of the stress-energy tensor. To examine these constraints for simple systems is to realize the beauty and power of the equation $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$.

§5.10. SAMPLE APPLICATIONS OF $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$

Newtonian fluid characterized by $|v^j| \ll 1$, $p \ll \rho$

The equation of motion $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ makes contact with the classical (Newtonian) equations of hydrodynamics, when applied to a nearly Newtonian fluid. Such a fluid has low velocities relative to the Lorentz frame used, $|v^j| \ll 1$; and in its rest frame its pressure is small compared to its density of mass-energy, $p/\rho = p/\rho c^2 \ll 1$. For example, the air in a hurricane has

$$|v^j| \sim 100 \text{ km/hour} \sim 3{,}000 \text{ cm/sec} \sim 10^{-7}\, c = 10^{-7} \ll 1,$$

$$\frac{p}{\rho} \sim \frac{1 \text{ atmosphere}}{10^{-3} \text{ g/cm}^3} \sim \frac{10^6 \text{ dynes/cm}^2}{10^{-3} \text{ g/cm}^3} = 10^9 \frac{\text{cm}^2}{\text{sec}^2} \sim 10^{-12}\, c^2 = 10^{-12} \ll 1.$$

Stress-energy tensor and equation of motion for a Newtonian fluid

The stress-energy tensor for such a fluid has components

$$T^{00} = (\rho + p)u^0u^0 - p \approx \rho, \tag{5.37a}$$

$$T^{0j} = T^{j0} = (\rho + p)u^0u^j \approx \rho v^j, \tag{5.37b}$$

$$T^{jk} = (\rho + p)u^ju^k + p\,\delta^{jk} \approx \rho v^j v^k + p\,\delta^{jk}; \tag{5.37c}$$

and the equation of motion $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ has components

$$T^{00}{}_{,0} + T^{0j}{}_{,j} = \partial\rho/\partial t + \nabla \cdot (\rho \boldsymbol{v}) = 0 \quad \text{("equation of continuity")}; \tag{5.38a}$$

and

$$T^{j0}{}_{,0} + T^{jk}{}_{,k} = \partial(\rho v^j)/\partial t + \partial(\rho v^j v^k)/\partial x^k + \partial p/\partial x^j = 0,$$

or, equivalently (by combining with the equation of continuity),

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\frac{1}{\rho}\nabla p \qquad \text{("Euler's equation")}. \tag{5.38b}$$

Box 5.5 derives and discusses these results from the Newtonian viewpoint.

Application of $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ to an electrically charged, vibrating rubber block

As a second application of $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$, consider a composite system: a block of rubber with electrically charged beads imbedded in it, interacting with an electromagnetic field. The block of rubber vibrates, and its accelerating beads radiate electromagnetic waves; at the same time, incoming electromagnetic waves push on the beads, altering the pattern of vibration of the block of rubber. The interactions shove 4-momentum back and forth between beaded block and electromagnetic field.

Box 5.5 NEWTONIAN HYDRODYNAMICS REVIEWED

Consider a classical, nonrelativistic, perfect fluid. Apply Newton's law $\boldsymbol{F} = m\boldsymbol{a}$ to a "fluid particle"; that is, to a small fixed mass of fluid followed in its progress through space:

$$\frac{d}{dt}(\text{momentum per unit mass}) = (\text{force per unit mass})$$

$$= \frac{(\text{force per unit volume})}{(\text{density})} = \frac{-(\text{gradient of pressure})}{(\text{density})}$$

or

$$\frac{d\boldsymbol{v}}{dt} = -\frac{1}{\rho}\nabla p. \tag{1}$$

Translate from time-rate of change following the fluid to time-rate of change as measured at a fixed location, finding

$$\begin{pmatrix}\text{rate of change}\\ \text{with time}\\ \text{following fluid}\end{pmatrix} = \begin{pmatrix}\text{rate of change}\\ \text{with time at}\\ \text{fixed location}\end{pmatrix} + \begin{pmatrix}\text{velocity}\\ \text{of fluid}\end{pmatrix} \cdot \begin{pmatrix}\text{rate of change}\\ \text{with position}\end{pmatrix}$$

or

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\frac{1}{\rho}\nabla p \tag{2}$$

or

$$\frac{\partial v_i}{\partial t} + v_{i,k}v_k = -\frac{1}{\rho}p_{,i}.$$

(Latin indices run from 1 to 3; summation convention; upper and lower indices used indifferently for space dimensions in flat space!) This is *Euler's fundamental equation* for the hydrodynamics of a perfect fluid.

Two further equations are needed to complete the description of a perfect fluid. One states the absence of heat transfer by requiring that the specific entropy (entropy per unit mass) be constant for each fluid "particle":

$$\frac{ds}{dt} = 0, \qquad \text{or} \qquad \frac{\partial s}{\partial t} + (\boldsymbol{v}\cdot\nabla)s = 0. \tag{3}$$

The final equation expresses the conservation of mass:

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{v}) = 0, \tag{4}$$

or

$$\frac{\partial \rho}{\partial t} + (\rho v_k)_{,k} = 0;$$

Box 5.5 (*continued*)

it is analagous in every way to the equation that expresses conservation of charge in electrodynamics and that bears the same name, "*equation of continuity*."

The Newtonian stress-energy tensor, like its relativistic counterpart, is linked to conservation of momentum and mass. Therefore examine the time-rate of change of the density of fluid momentum, ρv_i, contained in a unit volume; thus,

$$\partial(\rho v_i)/\partial t = -(\rho v_i v_k)_{,k} - p_{,i}. \tag{5}$$

Momentum flows into the little volume element on the left ("force equals time-rate of change of momentum") and out on the right; similarly at the other faces. Therefore the righthand side of (5) must represent the divergence of this momentum flux:

$$\partial(\rho v_i)/\partial t = -T_{ik,k}. \tag{6}$$

Consequently, we take for the momentum flux itself

$$T^{ik} = T_{ik} = \underbrace{\rho v_i v_k}_{\text{"convection"}} + \underbrace{\delta_{ik} p}_{\text{"push"}} . \tag{7}$$

For the momentum density, the Newtonian value is

$$T^{0i} = T^{i0} = \rho v_i. \tag{8}$$

With this notation, the equation for the time-rate of change of momentum becomes

$$\partial T^{i\mu}/\partial x^\mu = 0; \tag{9}$$

and with $T^{00} = \rho$, the equation of continuity reads

$$\partial T^{0\mu}/\partial x^\mu = 0. \tag{10}$$

In conclusion, these Newtonian considerations give a reasonable approximation to the relativistic stress-energy tensor:

$$\left\| \begin{array}{c:c} \rho & \rho v^j \\ \hdashline \rho v^i & \rho v^i v^j + \delta^{ij} p \end{array} \right\| \simeq \left\| \begin{array}{c:c} (p+\rho)u^0 u^0 - p & (p+\rho)u^0 u^j \\ \hdashline (p+\rho)u^0 u^i & (p+\rho)u^i u^j + \delta^{ij} p \end{array} \right\| \tag{11}$$

The 4-momentum of neither block nor field is conserved; neither $\boldsymbol{\nabla}\cdot\boldsymbol{T}_{\text{(block)}}$ nor $\boldsymbol{\nabla}\cdot\boldsymbol{T}_{\text{(em field)}}$ vanishes. But total 4-momentum must be conserved, so

$$\boldsymbol{\nabla}\cdot(\boldsymbol{T}_{\text{(block)}}+\boldsymbol{T}_{\text{(em field)}}) \text{ must vanish.} \tag{5.39}$$

For a general electromagnetic field interacting with any source, $\boldsymbol{\nabla}\cdot\boldsymbol{T}_{\text{(em field)}}$ has the form

$$T^{\mu\nu}_{\text{(em field)},\nu} = -F^{\mu\alpha}J_\alpha. \tag{5.40}$$

(This was derived in exercise 3.18 by combining $T^{\mu\nu}{}_{,\nu}=0$ with expression 5.22 for the electromagnetic stress-energy tensor, and with Maxwell's equations.) For our beaded block, $\boldsymbol{J}$ is the 4-current associated with the vibrating, charged beads, and $\boldsymbol{F}$ is the electromagnetic field tensor. The time component of equation (5.40) reads

$$\begin{aligned} T^{0\nu}_{\text{(em field)},\nu} &= -F^{0k}J_k = -\boldsymbol{E}\cdot\boldsymbol{J} \\ &= -\begin{pmatrix}\text{rate at which electric field } \boldsymbol{E} \text{ does work}\\ \text{on a unit volume of charged beads}\end{pmatrix}. \end{aligned} \tag{5.41}$$

For comparison, $T^{00}_{\text{(block)},0}$ is the rate at which the block's energy density changes with time, $-T^{0j}_{\text{(block)},j}$ is the contribution of the block's energy flux to this rate of change of energy density, and consequently their difference $T^{0\nu}_{\text{(block)},\nu}$ has the meaning

$$T^{0\nu}_{\text{(block)},\nu} = \begin{pmatrix}\text{rate at which mass-energy of block per}\\ \text{unit volume increases due to actions}\\ \text{other than internal mechanical forces}\\ \text{between one part of block and another}\end{pmatrix}. \tag{5.42}$$

Hence, the conservation law

$$(T^{0\nu}_{\text{(em field)}} + T^{0\nu}_{\text{(block)}})_{,\nu} = 0$$

says that the mass-energy of the block increases at precisely the same rate as the electric field does work on the beads. A similar result holds for momentum:

$$\begin{aligned} T^{k\nu}_{\text{(em field)},\nu}\boldsymbol{e}_k &= -F^{k\nu}J_\nu\boldsymbol{e}_k = -(J^0\boldsymbol{E}+\boldsymbol{J}\times\boldsymbol{B}) \\ &= -\begin{pmatrix}\text{Lorentz force per unit volume}\\ \text{acting on beads}\end{pmatrix}, \end{aligned} \tag{5.43}$$

$$T^{k\nu}_{\text{(block)},\nu}\boldsymbol{e}_k = \begin{pmatrix}\text{rate at which momentum per unit volume}\\ \text{of block increases due to actions}\\ \text{other than its own stresses}\end{pmatrix}; \tag{5.44}$$

so the conservation law

$$(T^{k\nu}_{\text{(em field)}} + T^{k\nu}_{\text{(block)}})_{,\nu} = 0$$

says that the rate of change of the momentum of the block equals the force of the electromagnetic field on its beads.

§5.11. ANGULAR MOMENTUM

Angular momentum defined and its integral conservation law derived

The symmetry, $T^{\mu\nu} = T^{\nu\mu}$, of the stress-energy tensor enables one to define a conserved angular momentum $J^{\alpha\beta}$, analogous to the linear momentum p^{α}. The angular momentum is defined relative to a specific but arbitrary origin—an event $\mathscr{A}$ with coordinates, in a particular Lorentz frame,

$$x^{\alpha}(\mathscr{A}) = a^{\alpha}. \tag{5.45}$$

The angular momentum about $\mathscr{A}$ is defined using the tensor

$$\mathscr{J}^{\alpha\beta\gamma} = (x^{\alpha} - a^{\alpha})T^{\beta\gamma} - (x^{\beta} - a^{\beta})T^{\alpha\gamma}. \tag{5.46}$$

(Note that $x^{\alpha} - a^{\alpha}$ is the vector separation of the "field point" x^{α} from the "origin" $\mathscr{A}$; $T^{\alpha\gamma}$ is here evaluated at the "field point".) Because of the symmetry of $\boldsymbol{T}$, $\mathscr{J}^{\alpha\beta\gamma}$ has vanishing divergence:

$$\begin{aligned}\mathscr{J}^{\alpha\beta\gamma}{}_{,\gamma} &= \delta^{\alpha}{}_{\gamma}T^{\beta\gamma} + (x^{\alpha} - a^{\alpha})\underbrace{T^{\beta\gamma}{}_{,\gamma}}_{0} - \delta^{\beta}{}_{\gamma}T^{\alpha\gamma} - (x^{\beta} - a^{\beta})\underbrace{T^{\alpha\gamma}{}_{,\gamma}}_{0} \\ &= T^{\beta\alpha} - T^{\alpha\beta} = 0.\end{aligned} \tag{5.47}$$

Consequently, its integral over any closed 3-surface vanishes

$$\oint_{\partial\mathscr{V}} \mathscr{J}^{\alpha\beta\gamma}\, d^3\Sigma_{\gamma} = 0 \tag{5.48}$$

(*"integral form of the law of conservation of angular momentum"*).

The integral over a spacelike surface of constant time t is

$$J^{\alpha\beta} = \int \mathscr{J}^{\alpha\beta 0}\, dx\, dy\, dz = \int [(x^{\alpha} - a^{\alpha})T^{\beta 0} - (x^{\beta} - a^{\beta})T^{\alpha 0}]\, dx\, dy\, dz. \tag{5.49}$$

Recalling that $T^{\beta 0}$ is momentum density, one sees that (5.49) has the same form as the equation "$\boldsymbol{J} = \boldsymbol{r} \times \boldsymbol{p}$" of Newtonian theory. Hence the name "total angular momentum" for $J^{\alpha\beta}$. Various aspects of this conserved angular momentum, including the tie to its Newtonian cousin, are explored in Box 5.6.

EXERCISES

Exercise 5.2. CHARGE CONSERVATION

Exercise 3.16 revealed that the charge-current 4-vector $\boldsymbol{J}$ satisfies the differential conservation law $\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$. Write down the corresponding integral conservation law, and interpret it for the four closed surfaces of Fig. 5.3.

Exercise 5.3. PARTICLE PRODUCTION

Inside highly evolved, massive stars, the temperature is so high that electron-positron pairs are continually produced and destroyed. Let $\boldsymbol{S}$ be the number-flux vector for electrons and positrons, and denote its divergence by

$$\epsilon \equiv \boldsymbol{\nabla} \cdot \boldsymbol{S}. \tag{5.50}$$

Box 5.6 ANGULAR MOMENTUM

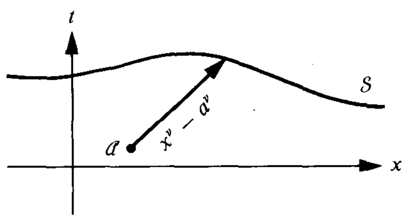
A. Definition of Angular Momentum

(a) Pick an arbitrary spacelike hypersurface $\mathcal{S}$ and an arbitrary event $\mathcal{A}$ with coordinates $x^\alpha(\mathcal{A}) \equiv a^\alpha$. (Use globally inertial coordinates throughout.)

(b) Define *"total angular momentum on $\mathcal{S}$ about $\mathcal{A}$"* to be

$$J^{\mu\nu} \equiv \int_{\mathcal{S}} \mathcal{J}^{\mu\nu\alpha}\, d^3\Sigma_\alpha,$$

$$\mathcal{J}^{\mu\nu\alpha} \equiv (x^\mu - a^\mu)T^{\nu\alpha} - (x^\nu - a^\nu)T^{\mu\alpha}.$$



(c) If $\mathcal{S}$ is a hypersurface of constant time t, this becomes

$$J^{\mu\nu} = \int \mathcal{J}^{\mu\nu 0}\, dx\, dy\, dz.$$

B. Conservation of Angular Momentum

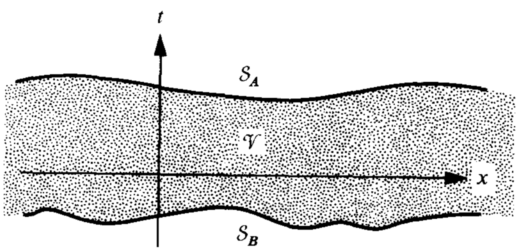
(a) $T^{\mu\nu}{}_{,\nu} = 0$ implies $\mathcal{J}^{\mu\nu\alpha}{}_{,\alpha} = 0$.

(b) This means that $J^{\mu\nu}$ *is independent of the hypersurface $\mathcal{S}$ on which it is calculated* (Gauss's theorem):

$$J^{\mu\nu}(\mathcal{S}_A) - J^{\mu\nu}(\mathcal{S}_B)$$

$$= \int_{\partial\mathcal{V}} \mathcal{J}^{\mu\nu\alpha}\, d^3\Sigma_\alpha$$

$$= \int_{\mathcal{V}} \mathcal{J}^{\mu\nu\alpha}{}_{,\alpha}\, d^4x = 0.$$



(Note: $\partial\mathcal{V} \equiv$ (boundary of $\mathcal{V}$) includes $\mathcal{S}_A$, $\mathcal{S}_B$, and timelike surfaces at spatial infinity; contribution of latter dropped—localized source.)

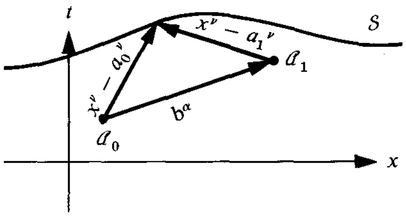
C. Change of Point About Which Angular Momentum is Calculated

Let b^α be vector from $\mathcal{A}_0$ to $\mathcal{A}_1$: $b^\alpha = a_1{}^\alpha - a_0{}^\alpha$. Then

$$J^{\mu\nu}(\text{about } \mathcal{A}_1) - J^{\mu\nu}(\text{about } \mathcal{A}_0)$$

$$= -b^\mu \int_{\mathcal{S}} T^{\nu\alpha}\, d^3\Sigma_\alpha + b^\nu \int_{\mathcal{S}} T^{\mu\alpha}\, d^3\Sigma_\alpha$$

$$= -b^\mu P^\nu + b^\nu P^\mu,$$



where P^μ is total 4-momentum.

Box 5.6 ***(continued)***

D. Intrinsic Angular Momentum

(a) Work, for a moment, in the system's rest frame, where

$$P^0 = M, \qquad P^j = 0, \qquad x_{CM}{}^j = \frac{1}{M}\int x^j T^{00}\, d^3x = \text{location of center of mass.}$$

Intrinsic angular momentum is defined as angular momentum about any event $(a^0, x_{CM}{}^j)$ on center of mass's world line. Its components are denoted $S^{\mu\nu}$ and work out to be

$$S^{0j} = 0,\ S^{jk} = \epsilon^{jk\ell} S^\ell,$$

where

$$\boldsymbol{S} \equiv \int (\boldsymbol{x} - \boldsymbol{x}_{CM}) \times (\textbf{momentum density})\, d^3x$$
$$\equiv \text{“intrinsic angular momentum vector.”}$$

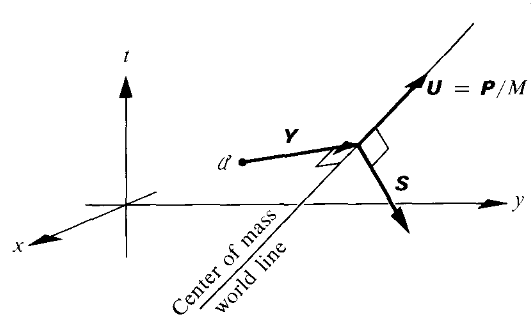
(b) Define "intrinsic angular momentum 4-vector" S^μ to be that 4-vector whose components in the rest frame are $(0, \boldsymbol{S})$; then the above equations say

$$S^{\mu\nu} = U_\alpha S_\beta \epsilon^{\alpha\beta\mu\nu},$$
$$U_\beta \equiv P_\beta / M = \text{4-velocity of center of mass,}$$
$$U_\beta S^\beta = 0.$$

E. Decomposition of Angular Momentum into Intrinsic and Orbital Parts

(a) Pick an arbitrary event $\mathscr{A}$, whose perpendicular displacement from center-of-mass world line is $-Y^\alpha$, so

$$U_\beta Y^\beta = 0.$$



(b) Then, by Part C, the angular momentum about $\mathscr{A}$ is

$$J^{\mu\nu} = \underbrace{U_\alpha S_\beta \epsilon^{\alpha\beta\mu\nu}}_{S^{\mu\nu}\text{ (intrinsic)}} + \underbrace{Y^\mu P^\nu - Y^\nu P^\mu}_{L^{\mu\nu}\text{ (orbital)}}.$$

(c) Knowing the angular momentum about $\mathscr{A}$, and the 4-momentum (and hence 4-velocity), one can calculate the vector from $\mathscr{A}$ to the center-of-mass world line,

$$Y^\mu = -J^{\mu\nu} P_\nu / M^2,$$

and the intrinsic angular momentum

$$S_\rho = \frac{1}{2} U^\sigma J^{\mu\nu} \epsilon_{\sigma\mu\nu\rho}.$$

Use Gauss's theorem to show that ϵ is the number of particles created (minus the number destroyed) in a unit four-dimensional volume of spacetime.

Exercise 5.4. INERTIAL MASS PER UNIT VOLUME

Consider a stressed medium in motion with ordinary velocity $|\boldsymbol{v}| \ll 1$ with respect to a specific Lorentz frame.

(a) Show by Lorentz transformations that the spatial components of the momentum density are

$$T^{0j} = \sum_k m^{jk} v^k, \tag{5.51}$$

where

$$m^{jk} = T^{\bar{0}\bar{0}}\,\delta^{jk} + T^{\bar{j}\bar{k}} \tag{5.52}$$

and $T^{\bar{\mu}\bar{\nu}}$ are the components of the stress-energy tensor in the rest frame of the medium. Throughout the solar system $T^{\bar{0}\bar{0}} \gg |T^{\bar{j}\bar{k}}|$ (see, e.g., discussion of hurricane in §5.10), so one is accustomed to write $T^{0j} = T^{\bar{0}\bar{0}} v^j$, i.e., "(momentum density) = (rest-mass density) $\times$ (velocity)". But inside a neutron star $T^{\bar{0}\bar{0}}$ may be of the same order of magnitude as $T^{\bar{j}\bar{k}}$, so one must replace "(momentum density) = (rest-mass density) $\times$ (velocity)" by equations (5.51) and (5.52), at low velocities.

(b) Derive equations (5.51) and (5.52) from Newtonian considerations plus the equivalence of mass and energy. (*Hint:* the total mass-energy carried past the observer by a volume V of the medium includes both the rest mass $T^{\bar{0}\bar{0}}V$ and the work done by forces acting across the volume's faces as they "push" the volume through a distance.)

(c) As a result of relation (5.51), the force per unit volume required to produce an acceleration dv^k/dt in a stressed medium, which is at rest with respect to the man who applies the force, is

$$F^j = dT^{0j}/dt = \sum_k m^{jk}\, dv^k/dt. \tag{5.53}$$

This equation suggests that one call m^{jk} the "inertial mass per unit volume" of a stressed medium at rest. In general m^{jk} is a symmetric 3-tensor. What does it become for the special case of a perfect fluid?

(d) Consider an isolated, stressed body at rest and in equilibrium ($T^{\alpha\beta}{}_{,0} = 0$) in the laboratory frame. Show that its total inertial mass, defined by

$$M^{ij} = \int_{\substack{\text{stressed}\\ \text{body}}} m^{ij}\, dx\, dy\, dz, \tag{5.54}$$

is isotropic and equals the rest mass of the body

$$M^{ij} = \delta^{ij} \int T^{00}\, dx\, dy\, dz. \tag{5.55}$$

Exercise 5.5. DETERMINANTS AND JACOBIANS

(a) Write out explicitly the sum defining d^2S_{01} in

$$d^2S_{\mu\nu} \equiv \epsilon_{\mu\nu\alpha\beta} \frac{\partial x^\alpha}{\partial a} \frac{\partial x^\beta}{\partial b}\, da\, db.$$

Thereby establish the formula

$$d^2S_{\mu\nu} = \epsilon_{\mu\nu|\alpha\beta|} \frac{\partial(x^\alpha, x^\beta)}{\partial(a, b)}\, da\, db = \frac{1}{2!} \epsilon_{\mu\nu\alpha\beta} \frac{\partial(x^\alpha, x^\beta)}{\partial(a, b)}\, da\, db.$$

(Expressions such as these should occur only under integral signs. In this exercise one may either supply an $\int \ldots$ wherever necessary, or else interpret the differentials in terms of the exterior calculus, $da\, db \longrightarrow \boldsymbol{da} \wedge \boldsymbol{db}$; see Box 5.4.) The notation used here for Jacobian determinants is

$$\frac{\partial(f, g)}{\partial(a, b)} = \begin{vmatrix} \dfrac{\partial f}{\partial a} & \dfrac{\partial f}{\partial b} \\[2ex] \dfrac{\partial g}{\partial a} & \dfrac{\partial g}{\partial b} \end{vmatrix}.$$

(b) By a similar inspection of a specific case, show that

$$d^3\Sigma_\mu \equiv \epsilon_{\mu\alpha\beta\gamma} \frac{\partial x^\alpha}{\partial a} \frac{\partial x^\beta}{\partial b} \frac{\partial x^\gamma}{\partial c}\, da\, db\, dc = \frac{1}{3!} \epsilon_{\mu\alpha\beta\gamma} \frac{\partial(x^\alpha, x^\beta, x^\gamma)}{\partial(a, b, c)}\, da\, db\, dc.$$

(c) Cite a precise definition of the value of a determinant as a sum of terms (with suitably alternating signs), with each term a product containing one factor from each row and simultaneously one factor from each column. Show that this definition can be stated (in the 4×4 case, with the $p \times p$ case an obvious extension) as

$$\det A \equiv \det \|A^\lambda{}_\rho\| = \epsilon_{\alpha\beta\gamma\delta} A^\alpha{}_0 A^\beta{}_1 A^\gamma{}_2 A^\delta{}_3.$$

(d) Show that

$$\det A = \frac{1}{4!} \delta^{\mu\nu\rho\sigma}_{\alpha\beta\gamma\delta} A^\alpha{}_\mu A^\beta{}_\nu A^\gamma{}_\rho A^\delta{}_\sigma$$

(for a definition of $\delta^{\mu\nu\rho\sigma}_{\alpha\beta\gamma\delta}$, see exercises 3.13 and 4.12).

(e) Use properties of the δ-symbol to show that the matrix A^{-1} inverse to A has entries $(A^{-1})^{\mu}{}_{\alpha}$ given by

$$(A^{-1})^{\mu}{}_{\alpha}(\det A) = \frac{1}{3!}\,\delta^{\mu\nu\rho\sigma}_{\alpha\beta\gamma\delta}A^{\beta}{}_{\nu}A^{\gamma}{}_{\rho}A^{\delta}{}_{\sigma}.$$

(f) By an "index-mechanics" computation, from the formula for det A in part (d) derive the following expression for the derivative of the logarithm of the determinant

$$\boldsymbol{d}\ln|\det A| = \operatorname{trace}(A^{-1}\,\boldsymbol{d}A).$$

Here $\boldsymbol{d}A$ is the matrix $\|\boldsymbol{d}A^{\alpha}{}_{\mu}\|$ whose entries are 1-forms.

Exercise 5.6. CENTROIDS AND SIZES

Consider an isolated system with stress-energy tensor $T^{\mu\nu}$, total 4-momentum P^{α}, magnitude of 4-momentum $M = (-\boldsymbol{P}\cdot\boldsymbol{P})^{1/2}$, intrinsic angular momentum tensor $S^{\alpha\beta}$, and intrinsic angular momentum vector S^{α}. (See Box 5.6.) An observer with 4-velocity u^{α} defines the *centroid* of the system, at his Lorentz time $x^0 = t$ and in his own Lorentz frame, by

$$X^{j}_{\boldsymbol{u}}(t) = (1/P^0)\int_{x^0=t} x^j T^{00}\,d^3x \qquad \text{in Lorentz frame where } \boldsymbol{u} = \partial\mathscr{P}/\partial x^0. \tag{5.56}$$

This centroid depends on (i) the particular system being studied, (ii) the 4-velocity $\boldsymbol{u}$ of the observer, and (iii) the time t at which the system is observed.

(a) Show that the centroid moves with a uniform velocity

$$dX^{j}_{\boldsymbol{u}}/dt = P^j/P^0, \tag{5.57}$$

corresponding to the 4-velocity

$$\boldsymbol{U} = \boldsymbol{P}/M. \tag{5.57$'$}$$

Note that this "4-velocity of centroid" is independent of the 4-velocity $\boldsymbol{u}$ used in defining the centroid.

(b) The centroid associated with the rest frame of the system (i.e., the centroid defined with $\boldsymbol{u} = \boldsymbol{U}$) is called the *center of mass;* see Box 5.6. Let $\boldsymbol{\xi}_{\boldsymbol{u}}$ be a vector reaching from any event on the center-of-mass world line to any event on the world line of the centroid associated with 4-velocity $\boldsymbol{u}$; thus the components of $\boldsymbol{\xi}_{\boldsymbol{u}}$ in any coordinate system are

$$\xi^{\alpha}_{\boldsymbol{u}} = X^{\alpha}_{\boldsymbol{u}} - X^{\alpha}_{\boldsymbol{U}}. \tag{5.58}$$

Show that $\boldsymbol{\xi}_{\boldsymbol{u}}$ satisfies the equation

$$[(\xi^{\alpha}_{\boldsymbol{u}}P^{\beta} - P^{\alpha}\xi^{\beta}_{\boldsymbol{u}}) - S^{\alpha\beta}]u_{\beta} = 0. \tag{5.59}$$

[*Hint:* perform the calculation in a Lorentz frame where $\boldsymbol{u} = \partial\mathscr{P}/\partial x^0$.]

(c) Show that, as seen in the rest-frame of the system at any given moment of time, the above equation reduces to the three-dimensional Euclidean equation

$$\boldsymbol{\xi}_{\boldsymbol{u}} = -(\boldsymbol{v}\times\boldsymbol{S})/M, \tag{5.59$'$}$$

where $\boldsymbol{v} = \boldsymbol{u}/u^0$ is the ordinary velocity of the frame associated with the centroid.

(d) Assume that the energy density measured by any observer anywhere in spacetime is non-negative ($\boldsymbol{u} \cdot \boldsymbol{T} \cdot \boldsymbol{u} \geq 0$ for all timelike $\boldsymbol{u}$). In the rest frame of the system, construct the smallest possible cylinder that is parallel to $\boldsymbol{S}$ and that contains the entire system ($T_{\alpha\beta} = 0$ everywhere outside the cylinder). Show that the radius r_0 of this cylinder is limited by

$$r_0 \geq |\boldsymbol{S}|/M. \tag{5.60}$$

Thus, a system with given intrinsic angular momentum $\boldsymbol{S}$ and given mass M has a minimum possible size $r_{0\,\mathrm{min}} = |\boldsymbol{S}|/M$ as measured in its rest frame.

CHAPTER 6

ACCELERATED OBSERVERS

The objective world simply is; *it does not* happen. *Only to the gaze of my consciousness, crawling upward along the life line* [*world line*] *of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time.*

HERMAN WEYL (1949, p. 116)

§6.1. ACCELERATED OBSERVERS CAN BE ANALYZED USING SPECIAL RELATIVITY

It helps in analyzing gravitation to consider a situation where gravity is mocked up by acceleration. Focus attention on a region so far from any attracting matter, and so free of disturbance, that (to some proposed degree of precision) spacetime there can be considered to be flat and to have Lorentz geometry. Let the observer acquire the feeling that he is subject to gravity, either because of jet rockets strapped to his legs or because he is in a rocket-driven spaceship. How does physics look to him?

Dare one answer this question? At this early stage in the book, is one not too ignorant of gravitation physics to predict what physical effects will be measured by an observer who thinks he is in a gravitational field, although he is really in an accelerated spaceship? Quite the contrary; special relativity was developed precisely to predict the physics of accelerated objects—e.g., the radiation from an accelerated charge. Even the fantastic accelerations

Accelerated motion and accelerated observers can be analyzed using special relativity

$$a_{\text{nuclear}} \sim v^2/r \sim 10^{31}\ \text{cm/sec}^2 \sim 10^{28}\ \text{“earth gravities”}$$

suffered by a neutron bound in a nucleus, and the even greater accelerations met in high-energy particle-scattering events, are routinely and accurately treated within

1

Box 6.1 GENERAL RELATIVITY IS BUILT ON SPECIAL RELATIVITY

A tourist in a powered interplanetary rocket feels "gravity." Can a physicist by local effects convince him that this "gravity" is bogus? Never, says Einstein's principle of the local equivalence of gravity and accelerations. But then the physicist will make no errors if he deludes himself into treating true gravity as a local illusion caused by acceleration. Under this delusion, he barges ahead and solves gravitational problems by using special relativity: if he is clever enough to divide every problem into a network of local questions, each solvable under such a delusion, then he can work out all influences of any gravitational field. Only three basic principles are invoked: special-relativity physics, the equivalence principle, and the local nature of physics. They are simple and clear. To apply them, however, imposes a double task: (1) take spacetime apart into locally flat pieces (where the principles are valid), and (2) put these pieces together again into a comprehensible picture. To undertake this dissection and reconstitution, to see curved dynamic spacetime inescapably take form, and to see the consequences for physics: that is general relativity.

the framework of special relativity. The theoretician who confidently applies special relativity to antiproton annihilations and strange-particle resonances is not about to be frightened off by the mere illusions of a rocket passenger who gullibly believed the travel brochures advertising "earth gravity all the way." When spacetime is flat, move however one will, special relativity can handle the job. (It can handle bigger jobs too; see Box 6.1.) The essential features of *how* special relativity handles the job are summarized in Box 6.2 for the benefit of the Track-1 reader, who can skip the rest of the chapter, and also for the benefit of the Track-2 reader, who will find it useful background for the rest of the chapter.

Box 6.2 ACCELERATED OBSERVERS IN BRIEF

An accelerated observer can carry clocks and measuring rods with him, and can use them to set up a reference frame (coordinate system) in his neighborhood.

His clocks, if carefully chosen so their structures are affected negligibly by acceleration (e.g., atomic clocks), will tick at the same rate as unaccelerated clocks moving momentarily along with him:

$$\Delta\tau \equiv \begin{pmatrix} \text{time interval ticked off} \\ \text{by observer's clocks as he} \\ \text{moves a vector displacement} \\ \boldsymbol{\xi} \text{ along his world line} \end{pmatrix} = [-\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\xi})]^{1/2}.$$

And his rods, if chosen to be sufficiently rigid, will measure the same lengths as momentarily comoving, unaccelerated rods do. (For further discussion, see §16.4, and Boxes 16.2 to 16.4.)

Let the observer's coordinate system be a Cartesian latticework of rods and clocks, with the origin of the lattice always on his world line. He must keep his latticework small:

$$l \equiv \begin{pmatrix}\text{spatial dimensions}\\ \text{of lattice}\end{pmatrix} \ll \begin{pmatrix}\text{the acceleration measured}\\ \text{by accelerometers he carries}\end{pmatrix}^{-1} \equiv \frac{1}{g}.$$

At distances l away from his world line, strange things of dimensionless magnitude gl happen to his lattice—e.g., the acceleration measured by accelerometers differs from g by a fractional amount $\sim gl$ (exercise 6.7); also, clocks initially synchronized with the clock on his world line get out of step (tick at different rates) by a fractional amount $\sim gl$ (exercise 6.6). (Note that an acceleration of one "earth gravity" corresponds to

$$g^{-1} \sim 10^{-3}\ \text{sec}^2/\text{cm} \sim 10^{18}\ \text{cm} \sim 1\ \text{light-year},$$

so the restriction $l \ll 1/g$ is normally not severe.)

To deduce the results of experiments and observations performed by an accelerated observer, one can analyze them in coordinate-independent, geometric terms, and then project the results onto the basis vectors of his accelerated frame. Alternatively, one can analyze the experiments and observations in a Lorentz frame, and then transform to the accelerated frame.

As deduced in this manner, the results of experiments performed locally (at $l \ll 1/g$) by an accelerated observer differ from the results of the same experiments performed in a Lorentz frame in only three ways:

(1) There are complicated fractional differences of order $gl \ll 1$ mentioned above, that can be made negligible by making the accelerated frame small enough.

(2) There are Coriolis forces of precisely the same type as are encountered in Newtonian theory (exercise 6.8). These the observer can get rid of by carefully preventing his latticework from rotating—e.g., by tying it to gyroscopes that he accelerates with himself by means of forces applied to their centers of mass (no torque!). Such a nonrotating latticework has "Fermi-Walker transported" basis vectors (§6.5),

$$\frac{d\boldsymbol{e}_{\alpha'}}{d\tau} = \boldsymbol{u}(\boldsymbol{a} \cdot \boldsymbol{e}_{\alpha'}) - \boldsymbol{a}(\boldsymbol{u} \cdot \boldsymbol{e}_{\alpha'}), \tag{1}$$

where $\boldsymbol{u}$ = 4-velocity, and $\boldsymbol{a} = d\boldsymbol{u}/d\tau$ = 4-acceleration.

(3) There are inertial forces of precisely the same type as are encountered in Newtonian theory (exercise 6.8). These are due to the observer's acceleration, and he cannot get rid of them except by stopping his accelerating.

2

1

2

§6.2. HYPERBOLIC MOTION

The rest of this chapter is Track 2.
It depends on no preceding Track-2 material.
It is needed as preparation for
(1) the mathematical analysis of gyroscopes in curved spacetime (exercise 19.2, §40.7), and
(2) the mathematical theory of the proper reference frame of an accelerated observer (§13.6).
It will be helpful in many applications of gravitation theory (Chapters 18–40).

Study a rocket passenger who feels "gravity" because he is being accelerated in flat spacetime. Begin by describing his motion relative to an inertial reference frame. His 4-velocity satisfies the condition $\boldsymbol{u}^2 = -1$. To say that it is fixed in magnitude is to say that the 4-acceleration,

$$\boldsymbol{a} = d\boldsymbol{u}/d\tau, \tag{6.1}$$

is orthogonal to the 4-velocity:

$$0 = (d/d\tau)(-1/2) = (d/d\tau)\left(\frac{1}{2}\boldsymbol{u}\cdot\boldsymbol{u}\right) = \boldsymbol{a}\cdot\boldsymbol{u}. \tag{6.2}$$

This equation implies that $a^0 = 0$ in the rest frame of the passenger (that Lorentz frame, where, at the instant in question, $\boldsymbol{u} = \boldsymbol{e}_0$); in this frame the space components of a^μ reduce to the ordinary definition of acceleration, $a^i = d^2x^i/dt^2$. From the components $a^\mu = (0; a^i)$ in the rest frame, then, one sees that the magnitude of the acceleration in the rest frame can be computed as the simple invariant

$$\boldsymbol{a}^2 = a^\mu a_\mu = (d^2\mathbf{x}/dt^2)^2_{\text{as measured in rest frame}}.$$

Consider, for simplicity, an observer who feels always a constant acceleration g. Take the acceleration to be in the x^1 direction of some inertial frame, and take $x^2 = x^3 = 0$. The equations for the motion of the observer in that inertial frame become

$$\frac{dt}{d\tau} = u^0, \qquad \frac{dx}{d\tau} = u^1; \qquad \frac{du^0}{d\tau} = a^0, \qquad \frac{du^1}{d\tau} = a^1. \tag{6.3}$$

Write out the three algebraic equations

$$u^\mu u_\mu = -1,$$
$$u^\mu a_\mu = -u^0a^0 + u^1a^1 = 0,$$

and

$$a^\mu a_\mu = g^2.$$

Solve for the acceleration, finding

$$a^0 = \frac{du^0}{d\tau} = gu^1, \qquad a^1 = \frac{du^1}{d\tau} = gu^0. \tag{6.4}$$

These linear differential equations can be solved immediately. The solution, with a suitable choice of the origin, reads

$$t = g^{-1}\sinh g\tau, \qquad x = g^{-1}\cosh g\tau. \tag{6.5}$$

Uniformly accelerated observer moves on hyperbola in spacetime diagram

Note that $x^2 - t^2 = g^{-2}$. The world line is a hyperbola in a spacetime diagram ("hyperbolic motion"; Figure 6.1). Several interesting aspects of this motion are

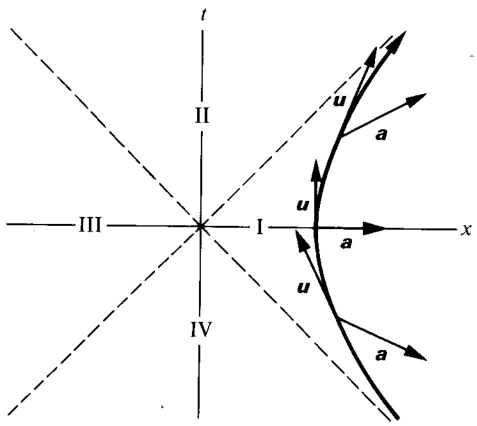


Figure 6.1.
Hyperbolic motion. World line of an object that (or an observer who) experiences always a fixed acceleration g with respect to an inertial frame that is instantaneously comoving (different inertial frames at different instants!). The 4-acceleration $\boldsymbol{a}$ is everywhere orthogonal (Lorentz geometry!) to the 4-velocity $\boldsymbol{u}$.

treated in the exercises. Let the magnitude of the constant acceleration g be the acceleration of gravity, $g = 980\ \text{cm/sec}^2$ experienced on earth: $g \simeq (10^3\ \text{cm/sec}^2)/(3 \times 10^{10}\ \text{cm/sec})^2 = (3 \times 10^7\ \text{sec} \cdot 3 \times 10^{10}\ \text{cm/sec})^{-1} = (1\ \text{light-year})^{-1}$. Thus the observer will attain relativistic velocities after maintaining this acceleration for something like one year of his own proper time. He can outrun a photon if he has a head start on it of one light-year or more.

EXERCISES

Exercise 6.1. A TRIP TO THE GALACTIC NUCLEUS

Compute the proper time required for the occupants of a rocket ship to travel the ~30,000 light-years from the Earth to the center of the Galaxy. Assume that they maintain an acceleration of one "earth gravity" ($10^3\ \text{cm/sec}^2$) for half the trip, and then decelerate at one earth gravity for the remaining half.

Exercise 6.2. ROCKET PAYLOAD

What fraction of the initial mass of the rocket can be payload for the journey considered in exercise 6.1? Assume an ideal rocket that converts rest mass into radiation and ejects all the radiation out the back of the rocket with 100 per cent efficiency and perfect collimation.

Exercise 6.3. TWIN PARADOX

(a) Show that, of all timelike world lines connecting two events $\mathscr{A}$ and $\mathscr{B}$, the one with the *longest* lapse of proper time is the unaccelerated one. (*Hint:* perform the calculation in the inertial frame of the unaccelerated world line.)

(b) One twin chooses to move from $\mathscr{A}$ to $\mathscr{B}$ along the unaccelerated world line. Show that the other twin, by an appropriate choice of accelerations, can get from $\mathscr{A}$ to $\mathscr{B}$ in arbitrarily small proper time.

(c) If the second twin prefers to ride in comfort, with the acceleration he feels never exceeding one earth gravity, g, what is the shortest proper time-lapse he can achieve between $\mathscr{A}$ and $\mathscr{B}$? Express the answer in terms of g and the proper time-lapse $\Delta\tau$ measured by the unaccelerated twin.

(d) Evaluate the answer numerically for several interesting trips.

Exercise 6.4. RADAR SPEED INDICATOR

A radar set measures velocity by emitting a signal at a standard frequency and comparing it with the frequency of the signal reflected back by another object. This redshift measurement is then converted, using the standard special-relativistic formula, into the corresponding velocity, and the radar reads out this velocity. How useful is this radar set as a velocity-measuring instrument for a uniformly accelerated observer?

(a) Consider this problem first for the special case where the object and the radar set are at rest with respect to each other at the instant the radar pulse is reflected. Compute the redshift $1 + z = \omega_e/\omega_0$ that the radar set measures in this case, and the resulting (incorrect) velocity it infers. Simplify by making use of the symmetries of the situation.

(b) Now consider the situation where the object has a non-zero velocity in the momentary rest frame of the observer at the instant it reflects the radar pulse. Compute the ratio of the actual relative velocity to the velocity read out by the radar set.

Exercise 6.5. RADAR DISTANCE INDICATOR

Use radar as a distance-measuring device. The radar set measures its proper time τ between the instant at which it emits a pulse and the later instant when it receives the reflected pulse. It then performs the simple computation $L_0 = \tau/2$ and supplies as output the "distance" L_0. How accurate is the output reading of the radar set for measuring the actual distance L to the object, when used by a uniformly accelerated observer? (L is defined as the distance in the momentary rest frame of the observer at the instant the pulse is reflected, which is at the observer's proper time halfway between emitting and receiving the pulse.) Give a correct formula relating $L_0 \equiv \tau/2$ to the actual distance L. Show that the reading L_0 becomes infinite as L approaches g^{-1}, where g is the observer's acceleration, as measured by an accelerometer he carries.

---

§6.3. CONSTRAINTS ON SIZE OF AN ACCELERATED FRAME

Difficulties in constructing "the coordinate system of an accelerated observer":

It is very easy to put together the words "the coordinate system of an accelerated observer," but it is much harder to find a concept these words might refer to. The most useful first remark one can make about these words is that, if taken seriously, they are self-contradictory. The definite article "the" in this phrase suggests that one is thinking of some unique coordinate system naturally associated with some specified accelerated observer, such as one whose world line is given in equation (6.5). If the coordinate system is indeed natural, one would expect that the coordinates of any event could be determined by a sufficiently ingenious observer by sending and receiving light signals. But from Figure 6.1 it is clear that the events composing one quarter of all spacetime (Zone III) can neither send light signals to, nor receive light signals from, the specified observer. Another half of spacetime suffers lesser disabilities in this respect: Zone II cannot send to the observer, Zone IV cannot receive from him. It is hard to see how the observer could define in any natural way a coordinate system covering events with which he has no causal relationship, which he cannot see, and from which he cannot be seen!

Breakdown in communication between observer and events at distance $l >$ (acceleration)$^{-1}$

Difficulties also occur when one considers an observer who begins at rest in one frame, is accelerated for a time, and maintains thereafter a constant velocity, at rest in some other inertial coordinate system. Do his motions define in any natural way

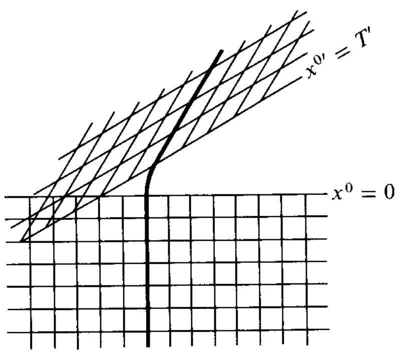


Figure 6.2.
World line of an observer who has undergone a brief period of acceleration. In each phase of motion at constant velocity, an inertial coordinate system can be set up. However, there is no way to reconcile these discordant coordinates in the region of overlap (beginning at distance g^{-1} to the left of the region of acceleration).

a coordinate system? Then this coordinate system (1) should be the inertial frame x^μ in which he was at rest for times x^0 less than 0, and (2) should be the other inertial frame $x^{\mu'}$ for times $x^{0'} > T'$ in which he was at rest in that other frame. Evidently some further thinking would be required to decide how to define the coordinates in the regions not determined by these two conditions (Figure 6.2). More serious, however, is the fact that these two conditions are inconsistent for a region of spacetime that satisfies simultaneously $x^0 < 0$ and $x^{0'} > T'$. In both examples of accelerated motion (Figures 6.1 and 6.2), the serious difficulties about defining a coordinate system begin only at a finite distance g^{-1} from the world line of the accelerated observer. The problem evidently has no solution for distances from the world line greater than g^{-1}. It does possess a natural solution in the immediate vicinity of the observer. This solution goes under the name of "Fermi-Walker transported orthonormal tetrad." The essential idea lends itself to simple illustration for hyperbolic motion, as follows.

Natural coordinates inconsistent at distance $l >$ (acceleration)$^{-1}$

§6.4. THE TETRAD CARRIED BY A UNIFORMLY ACCELERATED OBSERVER

An infinitesimal version of a coordinate system is supplied by a "tetrad," or "moving frame" (Cartan's "repère mobile"), or set of basis vectors $\boldsymbol{e}_{0'}$, $\boldsymbol{e}_{1'}$, $\boldsymbol{e}_{2'}$, $\boldsymbol{e}_{3'}$ (subscript tells which vector, not which component of one vector!) Let the time axis be the time axis of a comoving inertial frame in which the observer is momentarily at rest. Thus the zeroth basis vector is identical with his 4-velocity: $\boldsymbol{e}_{0'} = \boldsymbol{u}$. The space axes $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ are not affected by Lorentz transformations in the 1-direction. Therefore take $\boldsymbol{e}_{2'}$ and $\boldsymbol{e}_{3'}$ to be the unit basis vectors of the all-encompassing Lorentz frame relative to which the hyperbolic motion of the observer has already been described in equations (6.5): $\boldsymbol{e}_{2'} = \boldsymbol{e}_2$; $\boldsymbol{e}_{3'} = \boldsymbol{e}_3$. The remaining basis vector, $\boldsymbol{e}_{1'}$, orthogonal to the other three, is parallel to the acceleration vector, $\boldsymbol{e}_{1'} = g^{-1}\boldsymbol{a}$ [see equation (6.4)]. There is a more satisfactory way to characterize this moving frame: the time axis $\boldsymbol{e}_{0'}$ is the observer's 4-velocity, so he is always at rest in this frame; and the

Orthonormal tetrad of basis vectors carried by uniformly accelerated observer other three vectors $\boldsymbol{e}_{1'}$ are chosen in such a way as to be (1) orthogonal and (2) nonrotating. These basis vectors are:

$$
\begin{aligned}
(e_{0'})^\mu &= (\cosh g\tau;\ \sinh g\tau,\ 0,\ 0);\\
(e_{1'})^\mu &= (\sinh g\tau;\ \cosh g\tau,\ 0,\ 0);\\
(e_{2'})^\mu &= (0;\ 0,\ 1,\ 0);\\
(e_{3'})^\mu &= (0;\ 0,\ 0,\ 1).
\end{aligned}
\tag{6.6}
$$

There is a simple prescription to obtain these four basis vectors. Take the four basis vectors $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ of the original global Lorentz reference frame, and apply to them a simple boost in the 1-direction, of such a magnitude that $\boldsymbol{e}_{0'}$ comes into coincidence with the 4-velocity of the observer. The fact that these vectors are all orthogonal to each other and of unit magnitude is formally stated by the equation

$$\boldsymbol{e}_{\mu'} \cdot \boldsymbol{e}_{\nu'} = \eta_{\mu'\nu'}. \tag{6.7}$$

§6.5. THE TETRAD FERMI-WALKER TRANSPORTED BY AN OBSERVER WITH ARBITRARY ACCELERATION

Orthonormal tetrad of arbitrarily accelerated observer: should be "nonrotating"

Turn now from an observer, or an object, in hyperbolic motion to one whose acceleration, always finite, varies arbitrarily with time. Here also we impose three criteria on the moving, infinitesimal reference frame, or tetrad: (1) the basis vectors $\boldsymbol{e}_{\mu'}$ of the tetrad must remain orthonormal [equation (6.7)]; (2) the basis vectors must form a rest frame for the observer at each instant ($\boldsymbol{e}_{0'} = \boldsymbol{u}$); and (3) the tetrad should be "nonrotating."

This last criterion requires discussion. The basis vectors of the tetrad at any proper time τ must be related to the basis vectors $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ of some given inertial frame by a Lorentz transformation $\boldsymbol{e}_{\mu'}(\tau) = \Lambda^{\nu}{}_{\mu'}(\tau)\boldsymbol{e}_\nu$. Therefore the basis vectors at two successive instants must also be related to each other by a Lorentz transformation. But a Lorentz transformation can be thought of as a "rotation" in spacetime. The 4-velocity $\boldsymbol{u}$, always of unit magnitude, changes in direction. The very concept of acceleration therefore implies "rotation" of velocity 4-vector. How then is the requirement of "no rotation" to be interpreted? Demand that the tetrad $\boldsymbol{e}_{\mu'}(\tau)$ change from instant to instant by precisely that amount implied by the rate of change of $\boldsymbol{u} = \boldsymbol{e}_{0'}$, and by no additional arbitrary rotation. In other words, (1) accept the inevitable pseudorotation in the timelike plane defined by the velocity 4-vector and the acceleration, but (2) rule out any ordinary rotation of the three space vectors.

"Nonrotating" means rotation only in timelike plane of 4-velocity and 4-acceleration

Mathematics of rotation in 3-space

Nonrelativistic physics describes the rotation of a vector (components v_i) by an instantaneous angular velocity vector (components ω_i). This angular velocity appears in the formula for the rate of change of $\boldsymbol{v}$,

$$(dv_i/dt) = (\boldsymbol{\omega} \times \boldsymbol{v})_i = \varepsilon_{ijk}\omega_j v_k. \tag{6.8}$$

For the extension to four-dimensional spacetime, it is helpful to think of the rotation as occurring in the plane perpendicular to the angular velocity vector $\boldsymbol{\omega}$. Thus rewrite (6.8) as

$$dv_i/dt = -\Omega_{ik}v_k, \tag{6.9}$$

where

$$\Omega_{jk} = -\Omega_{kj} = \omega_i \varepsilon_{ijk} \tag{6.10}$$

has non-zero components only in the plane of the rotation. In other words, to speak of "a rotation in the (1, 2)-plane" is more useful than to speak of a rotation about the 3-axis. The concept of "plane of rotation" carries over to four dimensions. There a rotation in the (1, 2)-plane will leave constant not only the v_3 but also the v_0 component of the velocity. The four-dimensional definition of a rotation is

Mathematics of rotation in spacetime

$$\frac{dv^\mu}{d\tau} = -\Omega^{\mu\nu}v_\nu, \qquad \text{with} \qquad \Omega^{\mu\nu} = -\Omega^{\nu\mu}. \tag{6.11}$$

To test the appropriateness of this definition of a generalized rotation or infinitesimal Lorentz transformation, verify that it leaves invariant the length of the 4-vector:

$$d(v_\mu v^\mu)/d\tau = 2v_\mu(dv^\mu/d\tau) = -2\Omega^{\mu\nu}v_\mu v_\nu = 0. \tag{6.12}$$

The last expression vanishes because $\Omega^{\mu\nu}$ is antisymmetric, whereas $v_\mu v_\nu$ is symmetric. Note also that the antisymmetric tensor $\Omega^{\mu\nu}$ ("rotation matrix"; "infinitesimal Lorentz transformation") has $4 \times 3/2 = 6$ independent components. This number agrees with the number of components in a finite Lorentz transformation (three parameters for rotations, plus three parameters for the components of a boost). The "infinitesimal Lorentz transformation" here must (1) generate the appropriate Lorentz transformation in the timelike plane spanned by the 4-velocity and the 4-acceleration, and (2) exclude a rotation in any other plane, in particular, in any spacelike plane. The unique answer to these requirements is

$$\Omega^{\mu\nu} = a^\mu u^\nu - a^\nu u^\mu; \qquad \text{i.e., } \boldsymbol{\Omega} = \boldsymbol{a} \wedge \boldsymbol{u}. \tag{6.13}$$

Apply this rotation to a spacelike vector $\boldsymbol{w}$ orthogonal to $\boldsymbol{u}$ and $\boldsymbol{a}$, ($\boldsymbol{u} \cdot \boldsymbol{w} = 0$ and $\boldsymbol{a} \cdot \boldsymbol{w} = 0$). Immediately compute $\Omega^{\mu\nu}w_\nu = 0$. Thus verify the absence of any space rotation. Now check the over-all normalization of $\Omega^{\mu\nu}$ in equation (6.13). Apply the infinitesimal Lorentz transformation to the velocity 4-vector $\boldsymbol{u}$ of the observer. Thus insert $v^\mu = u^\mu$ in (6.11). It then reads

$$du^\mu/d\tau \equiv a^\mu = u^\mu(a^\nu u_\nu) - a^\mu(u^\nu u_\nu) = a^\mu.$$

This result is an identity, since $\boldsymbol{u} \cdot \boldsymbol{u} = -1$ and $\boldsymbol{u} \cdot \boldsymbol{a} = 0$.

A vector $\boldsymbol{v}$ that undergoes the indicated infinitesimal Lorentz transformation,

Fermi-Walker law of transport for "nonrotating" tetrad of basis vectors carried by an accelerated observer

$$dv^\mu/d\tau = (u^\mu a^\nu - u^\nu a^\mu)v_\nu, \tag{6.14}$$

is said to experience "Fermi-Walker transport" along the world line of the observer.

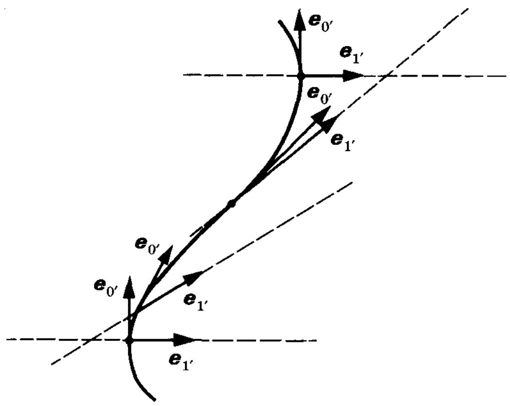


Figure 6.3.
Construction of spacelike hyperplanes (dashed) orthogonal to the world line (heavy line) of an accelerated particle at selected moments along that world line. Note crossing of hyperplanes at distance $g^{-1}(\tau)$ (time-dependent acceleration!) from the world line.

The natural moving frame associated with an accelerated observer consists of four orthonormal vectors, each of which is Fermi-Walker transported along the world line and one of which is $\boldsymbol{e}_{0'} = \boldsymbol{u}$ (the 4-velocity of the observer). Fermi-Walker transport of the space basis vectors $\boldsymbol{e}_{j'}$ can be achieved in practice by attaching them to gyroscopes (see Box 6.2 and exercise 6.9).

§6.6. THE LOCAL COORDINATE SYSTEM OF AN ACCELERATED OBSERVER

Tetrad used to construct "local coordinate system of accelerated observer"

Extend this moving frame or "infinitesimal coordinate system" to a "local coordinate system" covering a finite domain. Such local coordinates can escape none of the problems encountered in "hyperbolic motion" (Figure 6.1) and "briefly accelerated motion" (Figure 6.2). Therefore the local coordinate system has to be restricted to a region within a distance g^{-1} of the observer, where these problems do not arise. Figure 6.3 illustrates the construction of the local coordinates $\xi^{\mu'}$. At any given proper time τ the observer sits at a specific event $\mathscr{P}(\tau)$ along his world line. Let the displacement vector, from the origin of the original inertial frame to his position $\mathscr{P}(\tau)$, be $\boldsymbol{z}(\tau)$. At $\mathscr{P}(\tau)$ the observer has three spacelike basis vectors $\boldsymbol{e}_{1'}(\tau)$, $\boldsymbol{e}_{2'}(\tau)$, $\boldsymbol{e}_{3'}(\tau)$. The point $\mathscr{P}(\tau)$ plus those basis vectors define a spacelike hyperplane. The typical point of this hyperplane can be represented in the form

$$\begin{aligned}\boldsymbol{x} &= \xi^{1'}\boldsymbol{e}_{1'}(\tau) + \xi^{2'}\boldsymbol{e}_{2'}(\tau) + \xi^{3'}\boldsymbol{e}_{3'}(\tau) + \boldsymbol{z}(\tau) \\ &= \text{(separation vector from origin of original inertial frame).}\end{aligned} \tag{6.15}$$

Here the three numbers $\xi^{k'}$ play the role of Euclidean coordinates in the hyperplane. This hyperplane advances as proper time unrolls. Eventually the hyperplane cuts through the event $\mathscr{P}_0$ to which it is desired to assign coordinates. *Assign to this event as coordinates the numbers* $\xi^{0'} = \tau$, $\xi^{k'}$ *given by (6.15). Call these four numbers* *"coordinates relative to the accelerated observer."* In detail, the prescription for the determination of these four coordinates consists of the four equations

$$x^\mu = \xi^{k'}[e_{k'}(\tau)]^\mu + z^\mu(\tau), \tag{6.16}$$

in which the x^μ are considered as known, and the coordinates τ, $\xi^{k'}$ are considered unknowns.

At a certain distance from the accelerated world line, successive spacelike hyperplanes, instead of advancing with increasing τ, will be retrogressing. At this distance, and at greater distances, the concept of "coordinates relative to the accelerated observer" becomes ambiguous and has to be abandoned. To evaluate this distance, note that any sufficiently short section of the world line can be approximated by a hyperbola ("hyperbolic motion with acceleration g"), where the time-dependent acceleration $g(\tau)$ is given by the equation $g^2 = a^\mu a_\mu$.

Local coordinate system for uniformly accelerated observer

Apply the above general prescription to hyperbolic motion, arriving at the equations

$$\begin{aligned}
x^0 &= (g^{-1} + \xi^{1'})\sinh(g\xi^{0'}),\\
x^1 &= (g^{-1} + \xi^{1'})\cosh(g\xi^{0'}),\\
x^2 &= \xi^{2'},\\
x^3 &= \xi^{3'}.
\end{aligned} \tag{6.17}$$

The surfaces of constant $\xi^{0'}$ are the hyperplanes with $x^0/x^1 = \tanh g\xi^{0'}$ sketched in Figure 6.4. Substitute expressions (6.17) into the Minkowski formula for the line element to find

$$\begin{aligned}
ds^2 &= \eta_{\mu\nu}\, dx^\mu\, dx^\nu\\
&= -(1 + g\xi^{1'})^2(d\xi^{0'})^2 + (d\xi^{1'})^2 + (d\xi^{2'})^2 + (d\xi^{3'})^2.
\end{aligned} \tag{6.18}$$

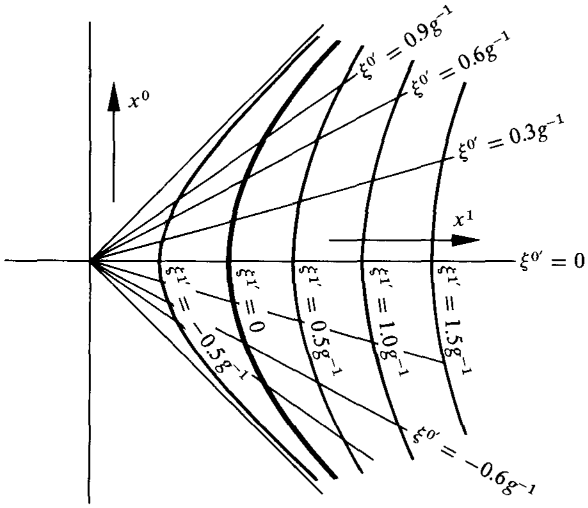


Figure 6.4.
Local coordinate system associated with an observer in hyperbolic motion (heavy black world line). The local coordinate system fails for $\xi^{1'}$ less than $-g^{-1}$.

The coefficients of $d\xi^{\mu'}\, d\xi^{\nu'}$ in this expansion are not the standard Lorentz metric components. The reason is clear. The $\xi^{\mu'}$ do not form an inertial coordinate system. However, at the position of the observer, $\xi^{1'} = 0$, the coefficients reduce to the standard form. Therefore these "local coordinates" approximate a Lorentz coordinate system in the immediate neighborhood of the observer.

EXERCISES

Exercise 6.6. CLOCK RATES VERSUS COORDINATE TIME IN ACCELERATED COORDINATES

Let a clock be attached to each grid point, $(\xi^{1'}, \xi^{2'}, \xi^{3'}) =$ constant, of the local coordinate system of an accelerated observer. Assume for simplicity that the observer is in hyperbolic motion. Use equation (6.18) to show that proper time as measured by a lattice clock differs from coordinate time at its lattice point:

$$d\tau/d\xi^{0'} = 1 + g\xi^{1'}.$$

(Of course, very near the observer, at $\xi^{1'} \ll g^{-1}$, the discrepancy is negligible.)

Exercise 6.7. ACCELERATION OF LATTICE POINTS IN ACCELERATED COORDINATES

Let an accelerometer be attached to each grid point of the local coordinates of an observer in hyperbolic motion. Calculate the magnitude of the acceleration measured by the accelerometer at $(\xi^{1'}, \xi^{2'}, \xi^{3'})$.

Exercise 6.8. OBSERVER WITH ROTATING TETRAD

An observer moving along an arbitrarily accelerated world line chooses *not* to Fermi-Walker transport his orthonormal tetrad. Instead, he allows it to rotate. The antisymmetric rotation tensor $\boldsymbol{\Omega}$ that enters into his transport law

$$d\boldsymbol{e}_{\alpha'}/d\tau = -\boldsymbol{\Omega}\cdot\boldsymbol{e}_{\alpha'} \tag{6.19}$$

splits into a Fermi-Walker part plus a spatial rotation part:

$$\Omega^{\mu\nu} = \underbrace{a^\mu u^\nu - a^\nu u^\mu}_{\Omega^{\mu\nu}_{(FW)}} + \underbrace{u_\alpha \omega_\beta \epsilon^{\alpha\beta\mu\nu}}_{\Omega^{\mu\nu}_{(SR)}} \tag{6.20}$$

$$\boldsymbol{\omega} = \text{a vector orthogonal to 4-velocity } \boldsymbol{u}.$$

(a) The observer chooses his time basis vector to be $\boldsymbol{e}_{0'} = \boldsymbol{u}$. Show that this choice is permitted by his transport law (6.19), (6.20).

(b) Show that $\Omega^{\mu\nu}_{(SR)}$ produces a rotation in the plane perpendicular to $\boldsymbol{u}$ and $\boldsymbol{\omega}$—i.e., that

$$\boldsymbol{\Omega}_{(SR)}\cdot\boldsymbol{u} = 0, \qquad \boldsymbol{\Omega}_{(SR)}\cdot\boldsymbol{\omega} = 0. \tag{6.21}$$

(c) Suppose the accelerated observer Fermi-Walker transports a second orthonormal tetrad $\boldsymbol{e}_{\alpha''}$. Show that the space vectors of his first tetrad rotate relative to those of his second tetrad with angular velocity vector equal to $\boldsymbol{\omega}$. *Hint:* At a moment when the tetrads coincide, show that (in three-dimensional notation, referring to the 3-space orthogonal to the observer's world line):

$$d(e_{j'} - e_{j''})/d\tau = \omega \times e_{j'}. \tag{6.22}$$

(d) The observer uses the same prescription [equation (6.16)] to set up local coordinates based on his rotating tetrad as for his Fermi-Walker tetrad. Pick an event $\mathscr{Q}$ on the observer's world line, set $\tau = 0$ there, and choose the original inertial frame of prescription (6.16) so (1) it comoves with the accelerated observer at $\mathscr{Q}$, (2) its origin is at $\mathscr{Q}$, and (3) its axes coincide with the accelerated axes at $\mathscr{Q}$. Show that these conditions translate into

$$z^{\mu}(0) = 0, \qquad \boldsymbol{e}_{\alpha'}(0) = \boldsymbol{e}_{\alpha}. \tag{6.23}$$

(e) Show that near $\mathscr{Q}$, equations (6.16) for the rotating, accelerated coordinates reduce to:

$$\begin{aligned} x^0 &= \xi^{0'} + a_{k'}\xi^{k'}\xi^{0'} + O([\xi^{\alpha'}]^3); \\ x^j &= \xi^{j'} + \frac{1}{2}a^j\xi^{0'2} + \epsilon^{jk\ell}\omega^k\xi^{\ell'}\xi^{0'} + O([\xi^{\alpha'}]^3). \end{aligned} \tag{6.24}$$

(f) A freely moving particle passes through the event $\mathscr{Q}$ with ordinary velocity $\boldsymbol{v}$ as measured in the inertial frame. By transforming its straight world line $x^j = v^j x^0$ to the accelerated, rotating coordinates, show that its coordinate velocity and acceleration there are:

$$\begin{aligned} (d\xi^{j'}/d\xi^{0'})_{\text{at}\,\mathscr{Q}} &= v^j; \\ (d^2\xi^{j'}/d\xi^{0'2})_{\text{at}\,\mathscr{Q}} &= \underbrace{-a^j}_{\substack{\text{inertial}\\ \text{acceleration}}} - \underbrace{2\epsilon^{jk\ell}\omega^k v^\ell}_{\substack{\text{Coriolis}\\ \text{acceleration}}} + \underbrace{2v^j a^k v^k}_{\substack{\text{relativisitc}\\ \text{correction to}\\ \text{inertial acceleration}}}. \end{aligned} \tag{6.25}$$

Exercise 6.9. THOMAS PRECESSION

Consider a spinning body (gyroscope, electron, . . .) that accelerates because forces act at its center of mass. Such forces produce no torque; so they leave the body's intrinsic angular-momentum vector $\boldsymbol{S}$ unchanged, except for the unique rotation in the $\boldsymbol{u} \wedge \boldsymbol{a}$ plane required to keep $\boldsymbol{S}$ orthogonal to the 4-velocity $\boldsymbol{u}$. Mathematically speaking, the body Fermi-Walker transports its angular momentum (no rotation in planes other than $\boldsymbol{u} \wedge \boldsymbol{a}$):

$$d\boldsymbol{S}/d\tau = (\boldsymbol{u} \wedge \boldsymbol{a}) \cdot \boldsymbol{S}. \tag{6.26}$$

This transport law applies to a spinning electron that moves in a circular orbit of radius r around an atomic nucleus. As seen in the laboratory frame, the electron moves in the x, y-plane with constant angular velocity, ω. At time $t = 0$, the electron is at $x = r$, $y = 0$; and its spin (as treated classically) has components

$$S^0 = 0, \qquad S^x = \frac{1}{\sqrt{2}}\hbar, \qquad S^y = 0, \qquad S^z = \frac{1}{2}\hbar.$$

Calculate the subsequent behavior of the spin as a function of laboratory time, $S^{\mu}(t)$. Answer:

$$\begin{aligned} S^x &= \frac{1}{\sqrt{2}}\hbar\,(\cos\omega t\cos\omega\gamma t + \gamma\sin\omega t\sin\omega\gamma t); \\ S^y &= \frac{1}{\sqrt{2}}\hbar\,(\sin\omega t\cos\omega\gamma t - \gamma\cos\omega t\sin\omega\gamma t); \\ S^z &= \frac{1}{2}\hbar; \qquad S^0 = -\frac{1}{\sqrt{2}}\hbar\, v\gamma\sin\omega\gamma t; \\ v &= \omega r; \qquad \gamma = (1 - v^2)^{-1/2}. \end{aligned} \tag{6.27}$$

Rewrite the time-dependent spatial part of this as

$$S^x + iS^y = \frac{\hbar}{\sqrt{2}}[e^{-i(\gamma-1)\omega t} + i(1-\gamma)\sin(\omega\gamma t)e^{i\omega t}]. \tag{6.28}$$

The first term rotates steadily in a retrograde direction with angular velocity

$$\begin{aligned}\omega_{\text{Thomas}} &= (\gamma - 1)\omega \\ &\approx \frac{1}{2}v^2\omega \ll \omega \text{ if } v \ll 1.\end{aligned} \tag{6.29}$$

It is called the Thomas precession. The second term rotates in a righthanded manner for part of an orbit ($0 < \omega\gamma t < \pi$) and in a lefthanded manner for the rest ($\pi < \omega\gamma t < 2\pi$). Averaged in time, it does nothing. Moreover, in an atom it is very small ($\gamma - 1 \ll 1$). It must be present, superimposed on the Thomas precession, in order to keep

$$\boldsymbol{S}\cdot\boldsymbol{u} = \boldsymbol{S}\cdot\boldsymbol{u} - S^0u^0 = 0, \tag{6.30}$$

and

$$\boldsymbol{S}^2 = \boldsymbol{S}^2 - (S^0)^2 = 3\hbar^2/4 = \text{constant}. \tag{6.31}$$

It comes into play with righthanded rotation when $\boldsymbol{S}\cdot\boldsymbol{u}$ is negative; it goes out of play when $\boldsymbol{S}\cdot\boldsymbol{u} = 0$; and it returns with lefthanded rotation when $\boldsymbol{S}\cdot\boldsymbol{u}$ turns positive.

The Thomas precession can be understood, alternatively, as a spatial rotation that results from the combination of successive boosts in slightly different directions. [See, e.g., exercise 103 of Taylor and Wheeler (1966).] For an alternative derivation of the Thomas precession (6.29) from "spinor formalism," see §41.4.

CHAPTER 7

INCOMPATIBILITY OF GRAVITY AND SPECIAL RELATIVITY

§7.1. ATTEMPTS TO INCORPORATE GRAVITY INTO SPECIAL RELATIVITY

This chapter is entirely Track 2.
It depends on no preceding Track-2 material.
It is not needed as preparation for any later chapter, but will be helpful in Chapter 18 (weak gravitational fields), and in Chapters 38 and 39 (experimental tests and other theories of gravity).

The discussion of special relativity so far has consistently assumed an absence of gravitational fields. Why must gravity be ignored in special relativity? This chapter describes the difficulties that gravitational fields cause in the foundations of special relativity. After meeting these difficulties, one can appreciate fully the curved-spacetime methods that Einstein introduced to overcome them.

Start, then, with what one already knows about gravity, Newton's formulation of its laws:

$$d^2x^i/dt^2 = -\partial\Phi/\partial x^i, \tag{7.1}$$

$$\nabla^2\Phi = 4\pi G\rho. \tag{7.2}$$

Newton's gravitational laws must be modified into four-dimensional, geometric form

These equations cannot be incorporated as they stand into special relativity. The equation of motion (7.1) for a particle is in three-dimensional rather than four-dimensional form; it requires modification into a four-dimensional vector equation for $d^2x^\mu/d\tau^2$. Likewise, the field equation (7.2) is not Lorentz-invariant, since the appearance of a three-dimensional Laplacian operator instead of a four-dimensional d'Alembertian operator means that the potential Φ responds instantaneously to changes in the density ρ at arbitrarily large distances away. In brief, Newtonian gravitational fields propagate with infinite velocity.

All straightforward modifications are unsatisfactory

One's first reaction to these problems might be to think that they are relatively straightforward to correct. Exercises at the end of this section study some relatively straightforward generalizations of these equations, in which the gravitational potential Φ is taken to be first a scalar, then a vector, and finally a symmetric tensor field. Each of these theories has significant shortcomings, and all fail to agree with observations. The best of them is the tensor theory (exercise 7.3, Box 7.1), which, however, Best modification (tensor theory in flat spacetime) is internally inconsistent; when repaired, it becomes general relativity.

is internally inconsistent and admits no exact solutions. This difficulty has been attacked in recent times by Gupta (1954, 1957, 1962), Kraichnan (1955), Thirring (1961), Feynman (1963), Weinberg (1965), Deser (1970). They show how the flat-space tensor theory may be modified within the spirit of present-day relativistic field theory to overcome these inconsistencies. By this field-theory route (part 5 of Box 17.2), they arrive uniquely at standard 1915 general relativity. Only at this end point does one finally recognize, from the mathematical form of the equations, that what ostensibly started out as a flat-space theory of gravity is really Einstein's theory, with gravitation being a manifestation of the curvature of spacetime. This book follows Einstein's line of reasoning because it keeps the physics to the fore.

EXERCISES

EXERCISES ON FLAT-SPACETIME THEORIES OF GRAVITY

The following three exercises provide a solid challenge. Happily, all three require similar techniques, and a solution to the most difficult one (exercise 7.3) is presented in Box 7.1. Therefore, it is reasonable to proceed as follows. (a) Work either exercise 7.1 (scalar theory of gravity) or 7.2 (vector theory of gravity), skimming exercise 7.3 and Box 7.1 (tensor theory of gravity) for outline and method, not for detail, whenever difficulties arise. (b) Become familiar with the results of the other exercise (7.2 or 7.1) by discussing it with someone who has worked it in detail. (c) Read in detail the solution to exercise 7.3 as presented in Box 7.1, and compare with the computed results for the other two theories. (d) Develop computational power by checking some detailed computations from Box 7.1.

Exercise 7.1. SCALAR GRAVITATIONAL FIELD, Φ

A. Consider the variational principle $\delta I = 0$, where

$$I = -m \int e^{\Phi} \left(-\eta_{\alpha\beta} \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda} \right)^{1/2} d\lambda, \tag{7.3}$$

Here m = (rest mass) and $z^\alpha(\lambda)$ = (parametrized world line) for a test particle in the scalar gravitational field Φ. By varying the particle's world line, derive differential equations governing the particle's motion. Write them using the particle's proper time as the path parameter,

$$d\tau = \left(-\eta_{\alpha\beta} \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda} \right)^{1/2} d\lambda,$$

so that $u^\alpha = dz^\alpha/d\tau$ satisfies $u^\alpha u^\beta \eta_{\alpha\beta} = -1$.

B. Obtain the field equation for $\Phi(\boldsymbol{x})$ implied by the variational principle $\delta I = 0$, where $I = \int \mathcal{L}\, d^4x$ and

$$\mathcal{L} = -\frac{1}{8\pi G} \eta^{\alpha\beta} \frac{\partial \Phi}{\partial x^\alpha} \frac{\partial \Phi}{\partial x^\beta} - \int m e^{\Phi}\, \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau. \tag{7.4}$$

Show that the second term here gives the same integral as that studied in part A (equation 7.3).

Discussion: The field equations obtained describe how a single particle of mass m generates the scalar field. If many particles are present, one includes in $\mathcal{L}$ a term $-\int m e^{\Phi}\, \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau$ for each particle.

C. Solve the field equation of part B, assuming a single source particle at rest. Also assume that $e^{\Phi} = 1$ is an adequate approximation in the neighborhood of the particle. Then check this assumption from your solution; i.e., what value does it assign to e^{Φ} at the surface of the earth? (Units with $c = 1$ are used throughout; one may also set $G = 1$, if one wishes.)

D. Now treat the static, spherically symmetric field Φ from part C as the field of the sun acting as a given external field in the variational principle of part A, and study the motion of a planet determined by this variational principle. Constants of motion are available from the spherical symmetry and time-independence of the integrand. Use spherical coordinates and assume motion in a plane. Derive a formula for the perihelion precession of a planet.

E. Pass to the limit of a zero rest-mass particle in the equations of motion of part A. Do this by using a parameter λ different from proper time, so chosen that $k^\mu = dx^\mu/d\lambda$ is the energy-momentum vector, and by taking the limit $m \longrightarrow 0$ with $k^0 = \gamma m = E$ remaining finite (so $u^0 = \gamma \longrightarrow \infty$). Use these equations to show that the quantities $q^\mu = k^\mu e^\Phi$ are constants of motion, and from this deduce that there is no bending of light by the sun in this scalar theory.

Exercise 7.2. VECTOR GRAVITATIONAL FIELD, Φ_μ

A. Verify that the variational principle $\delta I = 0$ gives Maxwell's equations by varying A_μ, and the Lorentz force law by varying $z^\mu(\tau)$, when

$$I = \frac{-1}{16\pi}\int F_{\mu\nu}F^{\mu\nu}\, d^4x + \frac{1}{2}m\int \frac{dz^\mu}{d\tau}\frac{dz_\mu}{d\tau}\, d\tau + e\int \frac{dz^\mu}{d\tau}A_\mu(z)\, d\tau. \tag{7.5}$$

Here $F_{\mu\nu}$ is an abbreviation for $A_{\nu,\mu} - A_{\mu,\nu}$. Hint: to vary $A_\mu(\boldsymbol{x})$, rewrite the last term as a spacetime integral by introducing a delta function $\delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]$ as in exercise 7.1, parts A and B.

B. Define, by analogy to the above, a vector gravitational field Φ_μ with $G_{\mu\nu} \equiv \Phi_{\nu,\mu} - \Phi_{\mu,\nu}$ using a variational principle with

$$I = +\frac{1}{16\pi G}\int G_{\mu\nu}G^{\mu\nu}\, d^4x + \frac{1}{2}m\int \frac{dz^\mu}{d\tau}\frac{dz_\mu}{d\tau}\, d\tau + m\int \Phi_\mu \frac{dz^\mu}{d\tau}\, d\tau. \tag{7.6}$$

(Note: if many particles are present, one must augment I by terms $\frac{1}{2}m\int(dz^\mu/d\tau)(dz_\mu/d\tau)\, d\tau + m\int\Phi_\mu(dz^\mu/d\tau)\, d\tau$ for each particle.) Find the "Coulomb" law in this theory, and verify that the coefficients of the terms in the variational principle have been chosen reasonably.

C. Compute the perihelion precession in this theory.

D. Compute the bending of light in this theory (i.e., scattering of a highly relativistic particle $u^0 = \gamma \longrightarrow \infty$), as it passes by the sun, because of the sun's Φ_μ field.

E. Obtain a formula for the total field energy corresponding to the Lagrangian implicit in part B. Use the standard method of Hamiltonian mechanics, with

$$I_{\text{field}} = \frac{1}{16\pi G}\int G_{\mu\nu}G^{\mu\nu}\, d^4x \equiv \int \mathcal{L}\, d^4x;$$

$\mathcal{L}$ is the Lagrangian density and $L \equiv \int\mathcal{L}\, d^3x$ is the Lagrangian. The corresponding Hamiltonian density ($\equiv$ energy density) is

$$\mathcal{H} = \sum_\mu \Phi_{\mu,0}\frac{\partial\mathcal{L}}{\partial\Phi_{\mu,0}} - \mathcal{L}.$$

Show that vector gravitational waves carry negative energy.

Exercise 7.3. SYMMETRIC TENSOR GRAVITATIONAL FIELD, $h_{\mu\nu} = h_{\nu\mu}$

Here the action principle is, as for the vector field, $\delta I = 0$, with $I = I_{\text{field}} + I_{\text{particle}} + I_{\text{interaction}}$. I_{particle} is the same as for the vector field:

$$I_{\text{particle}} = \frac{1}{2}m\int \frac{dz^\mu}{d\tau}\frac{dz_\mu}{d\tau}\, d\tau. \tag{7.7}$$

However, I_{field} and $I_{\text{interaction}}$ are different:

$$I_{\text{field}} = \int \mathcal{L}_f \, d^4x, \tag{7.8a}$$

$$\mathcal{L}_f = \frac{-1}{32\pi G}\left(\frac{1}{2}\, h_{\nu\beta,\alpha}\bar{h}^{\nu\beta,\alpha} - \bar{h}_{\mu\alpha}{}^{,\alpha}\bar{h}^{\mu\beta}{}_{,\beta}\right) \quad \begin{bmatrix}\text{Note that} \\ \text{one } h \text{ here} \\ \text{is not an } \bar{h}\end{bmatrix}, \tag{7.8b}$$

with

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h^{\sigma}{}_{\sigma}; \tag{7.8c}$$

$$I_{\text{interaction}} = \frac{1}{2}\int h_{\mu\nu}T^{\mu\nu}\, d^4x. \tag{7.9}$$

Here $T^{\mu\nu}$ is the stress-energy tensor for all nongravitational fields and matter present. For a system of point particles (used throughout this exercise),

$$T^{\mu\nu}(\boldsymbol{x}) = \int m \frac{dz^{\mu}}{d\tau}\frac{dz^{\nu}}{d\tau}\,\delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau. \tag{7.10}$$

A. Obtain the equations of motion of a particle by varying $z^{\mu}(\tau)$ in $\delta(I_{\text{particle}} + I_{\text{interaction}}) = 0$. Express your result in terms of the "gravitational force field"

$$\Gamma_{\nu\alpha\beta} = \frac{1}{2}(h_{\nu\alpha,\beta} + h_{\nu\beta,\alpha} - h_{\alpha\beta,\nu}) \tag{7.11}$$

derived from the tensor gravitational potentials $h_{\mu\nu} = h_{\nu\mu}$.

B. Obtain the field equations from $\delta(I_{\text{field}} + I_{\text{interaction}}) = 0$; express them in terms of

$$-H^{\mu\alpha\nu\beta} \equiv \bar{h}^{\mu\nu}\eta^{\alpha\beta} + \bar{h}^{\alpha\beta}\eta^{\mu\nu} - \bar{h}^{\alpha\nu}\eta^{\mu\beta} - \bar{h}^{\mu\beta}\eta^{\alpha\nu}. \tag{7.12}$$

Discuss gauge invariance, and the condition $\bar{h}^{\mu\alpha}{}_{,\alpha} = 0$.

C. Find the tensor gravitational potentials $h_{\mu\nu}$ due to the sun (treated as a point mass).

D. Compute the perihelion precession.

E. Compute the bending of light.

F. Consider a gravitational wave

$$\bar{h}^{\mu\nu} = A^{\mu\nu}\exp(ik_{\alpha}x^{\alpha}). \tag{7.13}$$

What conditions are imposed by the field equations? By the gauge condition

$$\bar{h}^{\mu\alpha}{}_{,\alpha} = 0? \tag{7.14}$$

Show that, by further gauge transformations

$$h_{\mu\nu} \longrightarrow h_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu} \tag{7.15}$$

that preserve the $\bar{h}^{\mu\alpha}{}_{,\alpha} = 0$ restrictions, further conditions

$$u_{\alpha}\bar{h}^{\alpha\mu} = 0, \qquad \bar{h}^{\alpha}{}_{\alpha} = 0 \tag{7.16}$$

can be imposed, where u^{α} is a fixed, timelike vector. It is sufficient to consider the case, obtained by a suitable choice of reference frame, where $u^{\alpha} = (1; 0, 0, 0)$ and $k^{\alpha} = (\omega; 0, 0, \omega)$.

G. From the Hamiltonian density

$$\mathcal{H} \equiv \dot{h}_{\mu\nu}(\partial\mathcal{L}/\partial\dot{h}_{\mu\nu}) - \mathcal{L} \tag{7.17}$$

for the field, show that the energy density of the waves considered in part F is positive.

H. Compute $T^{\mu\nu}{}_{,\nu}$ for the stress-energy tensor of particles $T^{\mu\nu}$ that appears in the action integral I. Does $T^{\mu\nu}{}_{,\nu}$ vanish (e.g., for the earth in orbit around the sun)? Why? Show that the coupled equations for fields and particles obtained from $\delta I = 0$ have no solutions.

(*continued on page 187*)

Box 7.1 AN ATTEMPT TO DESCRIBE GRAVITY BY A SYMMETRIC TENSOR FIELD IN FLAT SPACETIME [Solution to exercise 7.3]

Attempts to describe gravity within the framework of special relativity would naturally begin by considering the gravitational field to be a scalar (exercise 7.1) as it is in Newtonian theory, or a vector (exercise 7.2) by analogy to electromagnetism. Only after these are found to be deficient (e.g., no bending of light in either theory; negative-energy waves in the vector theory) would one face the computational complexities of a symmetric tensor gravitational potential, $h_{\mu\nu} = h_{\nu\mu}$, which has more indices.

The foundations of the most satisfactory of all tensor theories of gravity in flat spacetime are laid out at the beginning of exercise 7.3. The choice of the Lagrangian made there (equations 7.8) is dictated by the demand that $h_{\mu\nu}$ be a "Lorentz covariant, massless, spin-two field." The meaning of this demand, and the techniques of special relativity required to translate it into a set of field equations, are customarily found in books on elementary particle physics or quantum field theory; see, e.g., Wentzel (1949), Feynman (1963), or Gasiorowicz (1966). Fierz and Pauli (1939) were the first to write down this Lagrangian and investigate the resulting theory. The conclusions of the theory are spelled out here in the form of a solution to exercise 7.3.

A. Equation of Motion for a Test Particle (exercise 7.3A)

Carry out the integration in equation (7.9), using the particle stress-energy tensor of equation (7.10), to find

$$I_{p+i} \equiv I_{\text{particle}} + I_{\text{interaction}} = \frac{1}{2} m \int (\eta_{\mu\nu} + h_{\mu\nu}) \dot{z}^{\mu} \dot{z}^{\nu} \, d\tau, \tag{1}$$

where

$$\dot{z}^{\mu} \equiv dz^{\mu}/d\tau.$$

Then compute δI_{p+i}, and find that the coefficient of the arbitrary variation in path δz^{μ} vanishes if and only if

$$(d/d\tau)[(\eta_{\mu\nu} + h_{\mu\nu})\dot{z}^{\nu}] - \frac{1}{2} h_{\alpha\beta,\mu} \dot{z}^{\alpha} \dot{z}^{\beta} = 0.$$

Rewrite this equation of motion in the form

$$(\eta_{\mu\nu} + h_{\mu\nu})\ddot{z}^{\nu} + \Gamma_{\mu\alpha\beta} \dot{z}^{\alpha} \dot{z}^{\beta} = 0, \tag{2}$$

where $\Gamma_{\mu\alpha\beta}$ is defined in equation (7.11).

B$_1$. Field Equations (exercise 7.3B)

Use I_{field} and $I_{\text{interaction}}$ in the forms given in equations (7.8) and (7.9); but for the quickest and least messy derivation, do *not* use the standard Euler-Lagrange equations. Instead, compute directly the first-order change $\delta \mathcal{L}_f$ produced by a small

Box 7.1 (*continued*)

variation $\delta h_{\alpha\beta}$ of the field. For the second term of $\mathcal{L}_f$, it is clear (by relabeling dummy indices as needed) that varying each factor gives the same result, so the two terms from the product rule combine:

$$\delta(\bar{h}_{\mu\alpha}{}^{,\alpha}\, \bar{h}^{\mu\beta}{}_{,\beta}) = 2\bar{h}^{\mu\beta}{}_{,\beta}\, \delta\bar{h}_{\mu\alpha}{}^{,\alpha}.$$

A similar result holds for the first term of $\mathcal{L}_f$, in view of the identity $a_{\mu\nu}\bar{b}^{\mu\nu} = \bar{a}_{\mu\nu}b^{\mu\nu}$, which holds for the "bar" operation of equations (7.8); each side here is just $a_{\mu\nu}b^{\mu\nu} - \frac{1}{2}a^{\mu}{}_{\mu}b^{\nu}{}_{\nu}$. Consequently,

$$-(32\pi G)\,\delta\mathcal{L}_f = \bar{h}^{\nu\beta,\alpha}\,\delta h_{\nu\beta,\alpha} - 2\bar{h}^{\mu\beta}{}_{,\beta}\,\delta\bar{h}_{\mu\alpha}{}^{,\alpha}. \tag{3}$$

Next use this expression in δI_{field}; and, by an integration by parts, remove the derivatives from $\delta h_{\mu\nu}$, giving

$$\delta I_{\text{field}} = (32\pi G)^{-1}\int [\bar{h}^{\nu\beta,\alpha}{}_{,\alpha}\,\delta h_{\nu\beta} - 2\bar{h}^{\mu\beta}{}_{,\beta}{}^{,\alpha}\,\delta\bar{h}_{\mu\alpha}]\,d^4x.$$

To find the coefficient of $\delta h_{\mu\nu}$ in this expression, write (from equation 7.8c)

$$\delta\bar{h}_{\alpha\beta} = (\delta^{\mu}{}_{\alpha}\,\delta_{\nu\beta} - \frac{1}{2}\eta_{\alpha\beta}\eta^{\mu\nu})\,\delta h_{\mu\nu};$$

and then rearrange and relabel dummy (summation) indices to obtain

$$\delta I_{\text{field}} = (32\pi G)^{-1}\int [\bar{h}^{\mu\beta,\alpha}{}_{,\alpha}\,\delta h_{\nu\beta} - 2\bar{h}^{\mu\beta}{}_{,\beta}{}^{,\alpha}\,\delta\bar{h}_{\mu\alpha}]\,d^4x.$$

By combining this with $\delta I_{\text{interaction}} = \frac{1}{2}T^{\mu\nu}\,\delta h_{\mu\nu}\,d^4x$, and by using the symmetry $\delta h_{\mu\nu} = \delta h_{\nu\mu}$, obtain

$$-\bar{h}^{\mu\nu}{}_{,\alpha}{}^{,\alpha} - \eta^{\mu\nu}\bar{h}^{\alpha\beta}{}_{,\alpha\beta} + \bar{h}^{\mu\alpha}{}_{,\alpha}{}^{,\nu} + \bar{h}^{\nu\alpha}{}_{,\alpha}{}^{,\mu} = 16\pi G T^{\mu\nu}. \tag{4}$$

The definition made in equation (7.12) allows this to be rewritten as

$$H^{\mu\alpha\nu\beta}{}_{,\alpha\beta} = 16\pi G T^{\mu\nu}. \tag{4$'$}$$

B_2. Gauge Invariance (exercise 7.3B, continued)

The symmetries,

$$H^{\mu\alpha\nu\beta} = H^{[\mu\alpha][\nu\beta]} = H^{\nu\beta\mu\alpha},$$

of $H^{\mu\alpha\nu\beta}$ imply an identity

$$H^{\mu\alpha\nu\beta}{}_{,\alpha\beta\nu} = H^{\mu\alpha[\nu\beta]}{}_{,\alpha(\beta\nu)} \equiv 0$$

analogous to $F^{\mu\nu}{}_{,\nu\mu} \equiv 0$ in electromagnetism.

Thus $T^{\mu\nu}{}_{,\nu} = 0$ is required of the sources, just as is $J^{\mu}{}_{,\mu} = 0$ in electromagnetism (exercise 3.16). These identities make the field equations (4′) too weak to fix $h_{\mu\nu}$

completely. In particular, by direct substitution in equations (4), one verifies that to any solution one can add a gauge field

$$\begin{aligned} h_{\mu\nu}{}^{(\text{gauge})} &= \xi_{\mu,\nu} + \xi_{\nu,\mu}, \\ \bar{h}_{\mu\nu}{}^{(\text{gauge})} &= \xi_{\mu,\nu} + \xi_{\nu,\mu} - \eta_{\mu\nu}\xi^{\alpha}{}_{,\alpha}, \end{aligned} \tag{5}$$

without changing $T^{\mu\nu}$.

Let ξ_{μ} vanish outside some finite spacetime volume, but be otherwise arbitrary. Then $h_{\mu\nu}$ and $h_{\mu\nu} = h_{\mu\nu} + h_{\mu\nu}{}^{(\text{gauge})}$ both satisfy the source equation (4) for the same source $T^{\mu\nu}$ and the same boundary conditions at infinity. We therefore expect them to be physically equivalent.

By a specialization of the gauge analogous to the "Lorentz" specialization $A^{\alpha}{}_{,\alpha} = 0$ of electromagnetism (equation 3.58a; exercise 3.17), one imposes the condition

$$\bar{h}^{\mu\alpha}{}_{,\alpha} = 0. \tag{6}$$

This reduces the field equations (4) to the simple d'Alembertian form

$$\Box\bar{h}^{\mu\nu} \equiv \bar{h}^{\mu\nu}{}_{,\alpha}{}^{\alpha} = -16\pi T^{\mu\nu} \tag{7}$$

(see exercise 18.2). Here and henceforth we set $G = 1$ ("geometrized units").

C. Field of a Point Mass (exercise 7.3C)

For a static source, the wave equation (7) reduces to a Laplace equation

$$\nabla^2\bar{h}_{\mu\nu} = -16\pi T_{\mu\nu}.$$

The stress-energy tensor for a static point mass (equation 7.10) is $T^{00} = M\delta^3(x)$ and $T^{\mu k} = 0$. Put this into the Laplace equation, solve for $\bar{h}_{\mu\nu}$, and use equation (7.8c) to obtain $h_{\mu\nu}$. The result is:

$$h_{00} = 2M/r; \qquad h_{0k} = 0; \qquad h_{ik} = \delta_{ik}(2M/r) \tag{8}$$

(see equation 18.15a).

D. Perihelion Precession (exercise 7.3D)

Direct substitution of the potential (8) into the equations of motion (2) is tedious and not very instructive. Variational principles are popular in mechanics because they simplify such calculations. Return to the basic variational principle $\delta I_{p+i} = 0$ (equation 1), and insert the potential (8) for the sun. Convert to spherical coordinates so oriented that the orbit lies in the equatorial ($\theta = \pi/2$) plane:

$$I_{p+i} = \int L\, d\tau; \tag{9}$$

$$L = \frac{1}{2}m[-(1 - 2Mr^{-1})\dot{t}^2 + (1 + 2Mr^{-1})(\dot{r}^2 + r^2\dot{\phi}^2)]. \tag{10}$$

Box 7.1 (*continued*)

From the absence of explicit t-, ϕ-, and τ-dependence in L, infer three constants of motion: the canonical momenta

$$P_t \equiv -m\gamma = \partial L/\partial \dot{t}$$

(this defines γ) and

$$P_\phi \equiv m\alpha = \partial L/\partial \dot{\phi}$$

(this defines α); and the Hamiltonian

$$H = \dot{x}^\mu(\partial L/\partial \dot{x}^\mu) - L,$$

which can be set equal to $-\frac{1}{2}m$ by appropriate normalization of the path parameter τ. From these constants of the motion, derive an orbit equation as follows: (1) calculate $H = -\frac{1}{2}m$ in terms of r, $\dot{r}$, ϕ, and $\dot{t}$; (2) eliminate $\dot{t}$ and $\dot{\phi}$ in favor of the constants γ and α; (3) as in Newtonian orbit problems, define $u = M/r$, and write

$$\frac{du}{d\phi} = \frac{\dot{u}}{\dot{\phi}} = -\frac{M\dot{r}}{r^2\dot{\phi}} = -\frac{M}{\alpha}(1 + 2u)\dot{r};$$

(4) in H, eliminate $\dot{r}$ in favor of $du/d\phi$ via the above equation, and eliminate r in favor of u; (5) solve for $du/d\phi$. The result is

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = (\gamma^2 - 1 + 2u)\frac{M^2}{\alpha^2}\left[\frac{1 + 2u}{1 - 2u}\right]. \tag{11}$$

Neglecting cubic and higher powers of $u = GM/c^2r \sim (1 - \gamma^2)$ in this equation, derive the perihelion shift. (For details of method, see exercise 40.4, with the γ and α of this box renamed $\tilde{E}$ and $\tilde{L}$, and with the γ and β of that exercise set equal to 1 and 0.) The resulting shift per orbit is

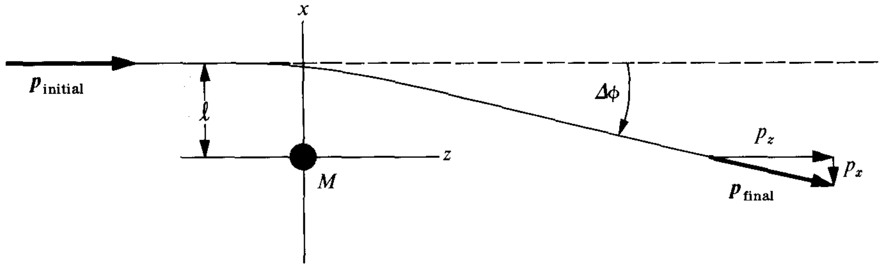
$$\Delta\phi = 8\pi M/r_0 + O([M/r_0]^2). \tag{12}$$

This is $\frac{4}{3}$ the prediction of general relativity; and it disagrees with the observations on Mercury (see Box 40.3).

E. Bending of Light (exercise 7.3E)

The deflection angle for light passing the sun is, on dimensional grounds, a small quantity, $\Delta\phi \sim M_\odot/R_\odot \sim 10^{-6}$; from the outset, one makes approximations based on this smallness. A diagram of the photon trajectory, set in the x, z-plane, shows that, for initial motion parallel to the z-axis, the deflection angle can be expressed in terms of the final momentum as $\Delta\phi = p_x/p_z$. Compute the final p_x by an integral along the trajectory,

$$p_x = \int_{-\infty}^{+\infty} (dp_x/dz)\, dz,$$



treating p_z as essentially constant. This computation requires generalization of the equation of motion (2) to the case of zero rest mass. To handle the limit $m \longrightarrow 0$, introduce a new parameter $\lambda = \tau/m$; then $p^\mu = m(dz^\mu/d\tau) = dz^\mu/d\lambda$. Also define $P_\mu = (\eta_{\mu\nu} + h_{\mu\nu})p^\nu$, since this quantity appears simply in equation (2) and agrees with p_μ in the limit $r \longrightarrow \infty$, where one will need to evaluate it. Then equation (2) reads, for any m, including $m = 0$,

$$\frac{dP_\mu}{d\lambda} = \frac{1}{2} h_{\alpha\beta,\mu} p^\alpha p^\beta.$$

On the righthand side here, since $h_{\alpha\beta,\mu}$ is small, a crude approximation to p^μ suffices; $p^1 = p^2 = 0$, $p^0 = p^3 = dz/d\lambda = \omega =$ constant. Thus,

$$\frac{dP_1}{d\lambda} = \frac{1}{2}(h_{00} + 2h_{03} + h_{33})_{,1}\omega^2$$

and

$$\frac{1}{p_3}\frac{dP_1}{dz} = \frac{1}{2}(h_{00} + 2h_{03} + h_{33})_{,1}.$$

For the sun, $h_{00} = h_{33} = 2M/r$, and $h_{03} = 0$ (equation 8), so

$$\Delta\phi = -\left(\frac{p_1}{p_3}\right)_{\text{final}} = -\left(\frac{P_1}{p_3}\right)_{\text{final}} = \int_{-\infty}^{\infty} \frac{2M\ell\, dz}{(\ell^2 + z^2)^{3/2}} = \frac{2M}{\ell}\int_{-\infty}^{\infty} \frac{d\zeta}{(1 + \zeta^2)^{3/2}} = \frac{4M}{\ell}. \quad (13)$$

For light grazing the sun, $\ell = R_\odot$, this gives $\Delta\phi = 4M_\odot/R_\odot$ radians $= 1''.75$, which is also the prediction of general relativity, and is consistent with the observations (see Box 40.1).

F. Gravitational Waves (exercise 7.3F)

The field equations (4) and gauge properties (5) of the present flat-spacetime theory are identical to those of Einstein's "linearized theory." Thus, the treatment of gravitational waves using linearized theory, which is presented in §§18.2, 35.3, and 35.4, applies here.

G. Positive Energy of the Waves (exercise 7.3G)

Computing a general formula for $\mathcal{H}$ from equation (7.17) is tedious, but it is sufficient to consider only the special case of a plane wave (equation 7.13)—or better still,

Box 7.1 ***(continued)***

a plane wave with only $h_{12} = h_{21} = f(z - t)$. Any gravitational wave can be constructed as a superposition of such plane waves. First compute the Langrangian for this case. According to equation (7.8), it reads

$$\mathcal{L}_f = (32\pi)^{-1}[(h_{12,0})^2 - (h_{12,3})^2].$$

Now the full content of the formula (7.17) defining $\mathcal{H}$ is precisely the following: start from the Lagrangian; keep all terms that are quadratic in time derivatives; omit all terms that are linear in time derivatives; and reverse the sign of terms that contain no time derivatives. The result is

$$\mathcal{H} = (32\pi)^{-1}[(h_{12,0})^2 + (h_{12,3})^2], \tag{14}$$

which is positive.

H. Self-Inconsistency of the Theory (exercise 7.3H)

From equation (7.10), find

$$T^{\mu\nu}{}_{,\nu} = m \int \frac{dz^\mu}{d\tau} \frac{dz^\nu}{d\tau} \frac{\partial}{\partial x^\nu} \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau.$$

But $\delta^4(\boldsymbol{x} - \boldsymbol{z})$ depends only on the difference $x^\mu - z^\mu$, so $-\partial/\partial z^\nu$ can replace $\partial/\partial x^\nu$ when acting on the δ-function. Noting that

$$\frac{dz^\nu}{d\tau} \frac{\partial}{\partial z^\nu} \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)] = \frac{d}{d\tau} \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)],$$

rewrite $T^{\mu\nu}{}_{,\nu}$ as

$$T^{\mu\nu}{}_{,\nu} = -m \int \dot{z}^\mu (d/d\tau)\, \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau = +m \int \ddot{z}^\mu\, \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau.$$

(The last step is obtained by an integration by parts.) Thus $T^{\mu\nu}{}_{,\nu} = 0$ holds *if and only if* $\ddot{z}^\mu = 0$. But $\ddot{z}^\mu = 0$ means that the gravitational fields have no effect on the motion of the particle. But this contradicts the equation of motion (2), which follows from the theory's variational principle. Thus, this tensor theory of gravity is inconsistent. [Stated briefly, equation (4) requires $T^{\mu\nu}{}_{,\nu} = 0$, while equation (2) excludes it.]

The fact that, in this theory, gravitating bodies cannot be affected by gravity, also holds for bodies made of arbitrary stress-energy (e.g., rubber balls or the Earth). Since all bodies gravitate, since the field equations imply $T^{\mu\nu}{}_{,\nu} = 0$, and since this "equation of motion for stress-energy" implies conservation of a body's total 4-momentum $P^\mu = \int T^{\mu 0}\, d^3x$, no body can be accelerated by gravity. The Earth cannot be attracted by the sun; it must fly off into interstellar space!

Straightforward steps to repair this inconsistency in the theory lead inexorably to general relativity (see Box 17.2 part 5). Having adopted general relativity as *the* theory of gravity, one can then use the present flat-spacetime theory as an approximation to it ("Linearized general relativity"; treated in Chapters 18, 19, and 35; see especially discussion at end of §18.3).

§7.2. GRAVITATIONAL RED SHIFT DERIVED FROM ENERGY CONSERVATION

Einstein argued against the existence of any ideal, straight-line reference system such as is assumed in Newtonian theory. He emphasized that nothing in a natural state of motion, not even a photon, could ever give evidence for the existence or location of such ideal straight lines.

Gravitational redshift derived from energy considerations

That a photon must be affected by a gravitational field Einstein (1911) showed from the law of conservation of energy, applied in the context of Newtonian gravitation theory. Let a particle of rest mass m start from rest in a gravitational field g at point $\mathscr{A}$ and fall freely for a distance h to point $\mathscr{B}$. It gains kinetic energy mgh. Its total energy, including rest mass, becomes

$$m + mgh. \tag{7.18}$$

Now let the particle undergo an annihilation at $\mathscr{B}$, converting its total rest mass plus kinetic energy into a photon of the same total energy. Let this photon travel upward in the gravitational field to $\mathscr{A}$. If it does not interact with gravity, it will have its original energy on arrival at $\mathscr{A}$. At this point it could be converted by a suitable apparatus into another particle of rest mass m (which could then repeat the whole process) plus an excess energy mgh that costs nothing to produce. To avoid this contradiction of the principal of conservation of energy, which can also be stated in purely classical terms, Einstein saw that the photon must suffer a red shift. The energy of the photon must decrease just as that of a particle does when it climbs out of the gravitational field. The photon energy at the top and the bottom of its path through the gravitational field must therefore be related by

$$E_{\text{bottom}} = E_{\text{top}}(1 + gh) = E_{\text{top}}(1 + g_{\text{conv}}h/c^2). \tag{7.19}$$

The drop in energy because of work done against gravitation implies a drop in frequency and an increase in wavelength (red shift; traditionally stated in terms of a red shift parameter, $z = \Delta\lambda/\lambda$); thus,

$$1 + z = \frac{\lambda_{\text{top}}}{\lambda_{\text{bottom}}} = \frac{h\nu_{\text{bottom}}}{h\nu_{\text{top}}} = \frac{E_{\text{bottom}}}{E_{\text{top}}} = 1 + gh. \tag{7.20}$$

The redshift predicted by this formula has been verified to 1 percent by Pound and Snider (1964, 1965), refining an experiment by Pound and Rebka (1960).

§7.3. GRAVITATIONAL REDSHIFT IMPLIES SPACETIME IS CURVED

An argument by Schild (1960, 1962, 1967) yields an important conclusion: the existence of the gravitational redshift shows that a consistent theory of gravity cannot be constructed within the framework of special relativity.

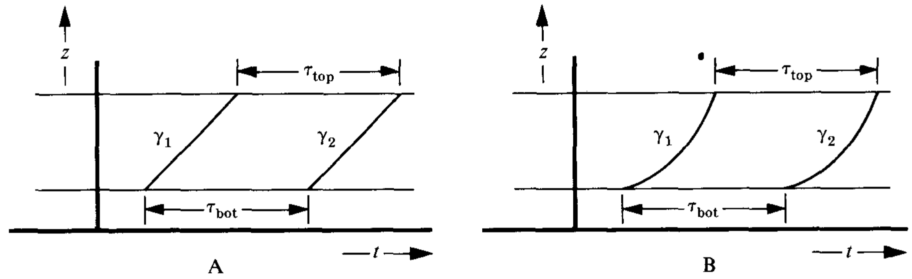


Figure 7.1.
Successive pulses of light rising from height z_1, to height $z_2 = z_1 + h$ against the gravitational field of the earth. The paths γ_1, and γ_2 must be exactly congruent, whether sloped at 45° (left) or having variable slope (right).

Assume gravity is described by an (unspecified) field in flat spacetime . . .

Whereas Einstein's argument (last section) was formulated in Newtonian theory, Schild's is formulated in special relativity. It analyzes gravitational redshift experiments in the field of the Earth, using a global Lorentz frame tied to the Earth's center. It makes no demand that free particles initially at rest remain at rest in this global Lorentz frame (except far from the Earth, where gravity is negligible). On the contrary, it demands that free particles be accelerated relative to the Lorentz frame by the Earth's gravitational field. It is indifferent to the mathematical nature of that field (scalar, vector, tensor, . . .), but it does insist that the gravitational accelerations agree with experiment. And, of course, it demands that proper lengths and times be governed by the metric of special relativity.

Schild's argument proceeds as follows. Consider one observer at rest on the Earth's surface at height z_1, and a second above the Earth's surface at height $z_2 = z_1 + h$ (Figure 7.1). The observers may verify that they are at rest relative to each other and relative to the Earth's Lorentz frame by, for instance, radar ranging to free particles that are at rest in the Earth's frame far outside its gravitational field. The bottom experimenter then emits an electromagnetic signal of a fixed standard frequency ω_b which is received by the observer on top. For definiteness, let the signal be a pulse exactly N cycles long. Then the interval of time* $\delta\tau_{bot}$ required for the emission of the pulse is given by $2\pi N = \omega_b\, \delta\tau_{bot}$. The observer at the top is then to receive these same N cycles of the electromagnetic wave pulse and measure the time interval* $\delta\tau_{top}$ required. By the definition of "frequency," it satisfies $2\pi N = \omega_t\, \delta\tau_{top}$. The redshift effect, established by experiment (for us) or by energy conservation (for Einstein), shows $\omega_t < \omega_b$; consequently the time intervals are different, $\delta\tau_{top} > \delta\tau_{bot}$. Transfer this information to the special-relativity spacetime diagram of the experiment (Figure 7.1). The waves are light rays; so one might show them as traveling along 45° null lines in the spacetime diagram (Figure 7.1,A). In this

*Proper time equals Lorentz coordinate time for both observers, since they are at rest in the Earth's Lorentz frame.

simplified but slightly inadequate form of the argument, one reaches a contradiction by noticing that here one has drawn a *parallelogram* in Minkowski spacetime in which two of the sides are unequal, $\tau_{\text{top}} > \tau_{\text{bot}}$, whereas a parallelogram in flat Minkowski spacetime cannot have opposite sides unequal. One concludes that *special relativity cannot be valid* over any sufficiently extended region. Globally, spacetime, as probed by the tracks of light rays and test particles, departs from flatness ("curvature"; Parts III and IV of this book), despite the fine fit that Lorentz-Minkowski flatness gives to physics locally.

This assumption is incompatible with gravitational redshift

Figure 7.1,B, repairs an oversimplification in this argument by recognizing that the propagation of light will be influenced by the gravitational field. Therefore photons might not follow straight lines in the diagram. Consequently, the world lines γ_1 and γ_2 of successive pulses are curves. However, the gravitational field is static and the experimenters do not move. Therefore nothing in the experimental setup changes with time. Whatever the path γ_1, the path γ_2 must be a *congruent* path of exactly the same shape, merely translated in time. On the basis of this congruence and the fact that the observers are moving on parallel world lines, one would again conclude, if flat Minkowski geometry were valid, that $\tau_{\text{bot}} = \tau_{\text{top}}$, thus contradicting the observed redshift experiment. The experimenters do not need to understand the propagation of light in a gravitational field. They need only use their radar apparatus to verify the fact that they are at rest relative to each other and relative to the source of the gravitational field. They know that, whatever influence the gravitational field has on their radar apparatus, it will not be a time-dependent influence. Moreover, they do not have to know how to compute their separation in order to verify that the separation remains constant. They only need to verify that the round-trip time required for radar pulses to go out to each other and back is the same every time they measure it.

Schild's redshift argument does not reveal what kind of curvature must exist, or whether the curvature exists in the neighborhood of the observational equipment or some distance away from it. It does say, however, quite unambigously, that the flat spacetime of special relativity is inadequate to describe the situation, and it should therefore motivate one to undertake the mathematical analysis of curvature in Part III.

Conclusion: spacetime is curved

§7.4. GRAVITATIONAL REDSHIFT AS EVIDENCE FOR THE PRINCIPLE OF EQUIVALENCE

Einstein (1908, 1911) elevated the idea of the universality of gravitational interactions to the status of a fundamental *principle of equivalence, that all effects of a uniform gravitational field are identical to the effects of a uniform acceleration of the coordinate system.* This principle generalized a result of Newtonian gravitation theory, in which a uniform acceleration of the coordinate system in equation (7.1) gives rises to a

Principle of equivalence: a uniform gravitational field is indistinguishable from a uniform acceleration of a reference frame supplementary uniform gravitational field. However, the Newtonian theory only gives this result for particle mechanics. Einstein's principle of equivalence asserts that a similar correspondence will hold for all the laws of physics, including Maxwell's equations for the electromagnetic field.

The rules of the game—the "scientific method"—require that experimental support be sought for any new theory or principle, and Einstein could treat the gravitational redshift as the equivalent of experimental confirmation of his principle of equivalence. There are two steps in such a confirmation: first, the theory or principle must predict an effect (the next paragraph describes how the equivalence principle implies the redshift); second, the predicted effect must be observed. With the Pound-Rebka-Snider experiments, one is in much better shape today than Einstein was for this second step. Einstein himself had to rely on the experiments supporting the general concept of energy conservation, plus the necessity of a redshift to preserve energy conservation, as a substitute for direct experimental confirmation.

Gravitational redshift derived from equivalence principle

The existence of the gravitational redshift can be deduced from the equivalence principle by considering two experimenters in a rocket ship that maintains a constant acceleration g. Let the distance between the two observers be h in the direction of the acceleration. Suppose for definiteness that the rocket ship was at rest in some inertial coordinate system when the bottom observer sent off a photon. It will require time $t = h$ for the photon to reach the upper observer. In that time the top observer acquires a velocity $v = gt = gh$. He will therefore detect the photon and observe a Doppler redshift $z = v = gh$. The results here are therefore identical to equation (7.20). The principle of equivalence of course requires that, if this redshift is observed in an experiment performed under conditions of uniform acceleration in the absence of gravitational fields, then the same redshift must be observed by an experiment performed under conditions where there is a uniform gravitational field, but no acceleration. Consequently, by the principle of equivalence, one can derive equation (7.20) as applied to the gravitational situation.

§7.5. LOCAL FLATNESS, GLOBAL CURVATURE

The equivalence principle helps one to discern the nature of the spacetime curvature, whose existence was inferred from Schild's argument. Physics is the same in an accelerated frame as it is in a laboratory tied to the Earth's surface. Thus, an Earth-bound lab can be regarded as accelerating upward, with acceleration g, relative to the Lorentz frames in its neighborhood.* Equivalently, relative to the lab and the Earth's surface, all Lorentz frames must accelerate downward. But the downward (radial) direction is different at different latitudes and longitudes. Hence, local Lorentz frames, initially at rest with respect to each other but on opposite sides of the Earth, subsequently fall toward the center and go flying through each other. Clearly they cannot be meshed to form the single global Lorentz frame, tied to the

Equivalence principle implies nonmeshing of local Lorentz frames near Earth (spacetime curvature!)

*This upward acceleration of the laboratory, plus equation (6.18) for the line element in an accelerated coordinate system, explains the nonequality of the bottom and top edges of the parallelograms in Figure 7.1.

Earth, that was assumed in Schild's argument. This nonmeshing of local Lorentz frames, like the nonmeshing of local Cartesian coordinates on a curved 2-surface, is a clear manifestation of spacetime curvature.

Geographers have similar problems when mapping the surface of the earth. Over small areas, a township or a county, it is easy to use a standard rectangular coordinate system. However, when two fairly large regions are mapped, each with one coordinate axis pointing north, then one finds that the edges of the maps overlap each other best if placed at a slight angle (spacetime analog: relative velocity of two local Lorentz frames meeting at center of Earth). It is much easier to start from a picture of a spherical globe, and then talk about how small flat maps might be used as good approximations to parts of it, than to start with a huge collection of small maps and try to piece them together to build up a picture of the globe. The exposition of the geometry of spacetime in this book will therefore take the first approach. Now that one recognizes that the problem is to fit together local, flat spacetime descriptions of physics into an over-all view of the universe, one should be happy to jump, in the next chapter, into a broadscale study of geometry. From this more advantageous viewpoint, one can then face the problem of discussing the relationship between the local inertial coordinate systems appropriate to two nearby regions that have slightly different gravitational fields.

Nonmeshing of local Lorentz frames motivates study of geometry

There are actually two distinguishable ways in which geometry enters the theory of general relativity. One is the geometry of lengths and angles in four-dimensional spacetime, which is inherited from the metric structure ds^2 of special relativity. Schild's argument already shows (without a direct appeal to the equivalence principle) that the special-relativistic ideas of length and angle must be modified. The modified ideas of metric structure lead to Riemannian geometry, which will be treated in Chapters 8 and 13. However, geometry also enters general relativity because of the equivalence principle. An equivalence principle can already be stated within Newtonian gravitational theory, in which no concepts of a *spacetime* metric enter, but only the Euclidean metric structure of three-dimensional *space*. The equivalence-principle view of Newtonian theory again insists that the local standard of reference be the freely falling particles. This requirement leads to the study of a *spacetime* geometry in which the curved world lines of freely falling particles are defined to be locally straight. They play the role in a curved spacetime geometry that straight lines play in flat spacetime. This "affine geometry" will be studied in Chapters 10–12. It leads to a quantitative formulation of the ideas of "covariant derivative" and "curvature" and even "curvature of Newtonian spacetime"!

Two types of geometry relevant to spacetime:

Riemannian geometry (lengths and angles)

Affine geometry ("straight lines" and curvature)

PART **III**

THE MATHEMATICS OF CURVED SPACETIME

Wherein the reader is exposed to the charms of a new temptress—
Modern Differential Geometry—and makes a decision:
to embrace her for eight full chapters; or,
having drunk his fill, to escape after one.

CHAPTER 8

DIFFERENTIAL GEOMETRY: AN OVERVIEW

I am coming more and more to the conviction that the necessity of our geometry cannot be demonstrated, at least neither by, nor for, the human intellect. . . . geometry should be ranked, not with arithmetic, which is purely aprioristic, but with mechanics. (*1817*)

We must confess in all humility that, while number is a product of our mind alone, space has a reality beyond the mind whose rules we cannot completely prescribe. (*1830*)

CARL FRIEDRICH GAUSS

§8.1. AN OVERVIEW OF PART III

Gravitation is a manifestation of spacetime curvature, and that curvature shows up in the deviation of one geodesic from a nearby geodesic ("relative acceleration of test particles"). The central issue of this part of the book is clear: *How can one quantify the "separation," and the "rate of change" of "separation," of two "geodesics" in "curved" spacetime?* A clear, precise answer requires new concepts.

Concepts to be developed in Part III:

"Separation" between geodesics will mean "vector." But the concept of vector as employed in flat Lorentz spacetime (a bilocal object: point for head and point for tail) must be sharpened up into the local concept of *tangent vector,* when one passes to curved spacetime. Chapter 9 does the sharpening. It also reveals how the passage to curved spacetime affects 1-forms and tensors.

Tangent vector

It takes one tool (vectors in curved geometry, Chapter 9) to define "separation" clearly as a vector; it takes another tool (parallel transport in curved spacetime, Chapter 10) to compare separation vectors at neighboring points and to define the "rate of change of separation." No transport, no comparison; no comparison, no meaning to the term "rate of change"! The notion of parallel transport founds itself

Geodesic

Covariant derivative

on the idea of *"geodesic,"* the world line of a freely falling particle. The special mathematical properties of a geodesic are explored in Chapter 10. That chapter uses geodesics to define parallel transport, uses parallel transport to define *covariant derivative,* and—completing the circle—uses covariant derivative to describe geodesics.

Geodesic deviation

Spacetime curvature

Chapter 11 faces up to the central issue: *geodesic deviation* ("rate of change of separation vector between two geodesics"), and its use in defining the *curvature* of spacetime.

But to define curvature is not enough. The man who would understand gravity deeply must also see curvature at work, producing relative accelerations of particles in Newtonian spacetime (Chapter 12); he must learn how, in Einstein spacetime, distances (metric) determine completely the curvature and the law of parallel transport (Chapter 13); he must be the master of powerful tools for computing curvature (Chapter 14); and he must grasp the geometric significance of the algebraic and differential symmetries of curvature (Chapter 15).

This chapter: a Track-1 overview of differential geometry

Unfortunately, such deep understanding requires time—far more time than one can afford in a ten-week or fifteen-week course, far more than a lone reader may wish to spend on first passage through the book. For the man who must rush on rapidly, this chapter contains a "Track-1" overview of the essential mathematical tools (§§8.4–8.7). From it one can gain an adequate, but not deep, understanding of spacetime curvature, of tidal gravitational forces, and of the mathematics of curved spacetime. This overview is also intended for the Track-2 reader; it will give him a taste of what is to come. The ambitious reader may also wish to consult other introductions to differential geometry (see Box 8.1).

Box 8.1 BOOKS ON DIFFERENTIAL GEOMETRY

There are several mathematics texts that may be consulted for a more detailed and extensive discussion of modern differential geometry along the line taken here. Bishop and Goldberg (1968) is the no. 1 reference. Hicks (1965) could be chosen as a current standard graduate-level text, with O'Neill (1966) at the undergraduate level introducing many of the same topics without presuming that the reader finds easy and obvious the current style in which pure mathematicians think and write. Auslander and MacKenzie (1963) at a somewhat more advanced level also allow for the reader to whom differential equations are more familiar than homomorphisms. Willmore (1959) is easy to read but presents no challenge, and leads to little progress in adapting to the style of current mathematics. Trautman (1965) and Misner (1964a, 1969a) are introductions somewhat similar to ours, except for deemphasis of pictures; like ours, they are aimed at the student of relativity. Flanders (1963) is easy and useful as an introduction to exterior differential forms; it also gives examples of their application to a wide variety of topics in physics and engineering.

§8.2. TRACK 1 VERSUS TRACK 2: DIFFERENCE IN OUTLOOK AND POWER

Nothing is more wonderful about the relation between Einstein's theory of gravity and Newton's theory than this, as discovered by Élie Cartan (1923, 1924): that both theories lend themselves to description in terms of curvature; that in both this curvature is governed by the density of mass-energy; and that this curvature allows itself to be defined and measured without any use of or reference to any concept of metric. The difference between the two theories shows itself up in this: Einstein's theory in the end (or in the beginning, depending upon how one presents it!) does define an interval between every event and every nearby event; Newton's theory not only does not, but even says that any attempt to talk of *spacetime* intervals violates Newton's laws. This being the case, Track 2 will forego for a time (Chapters 9–12) any use of a spacetime metric ("Einstein interval"). It will extract everything possible from a metric-free description of spacetime curvature (all of Newton's theory; important parts of Einstein's theory).

Preview of Track-2 differential geometry

Geodesic deviation is a measurer and definer of curvature, but the onlooker is forbidden to reduce a vector description of separation to a numerical measure of distance (no metric at this stage of the analysis): what an impossible situation! Nevertheless, that is exactly the situation with which Chapters 9–12 will concern themselves: how to do geometry without a metric. Speaking physically, one will overlook at this stage the fact that the geometry of the physical world is always and everywhere locally Lorentz, and endowed with a light cone, but one will exploit to the fullest the Galileo-Einstein principle of equivalence: in any given locality one can find a frame of reference in which every neutral test particle, whatever its velocity, is free of acceleration. The tracks of these neutral test particles define the geodesics of the geometry. These geodesics provide tools with which one can do much: define parallel transport (Chapter 10), define covariant derivative (Chapter 10), quantify geodesic deviation (Chapter 11), define spacetime curvature (Chapter 11), and explore Newtonian gravity (Chapter 12). Only after this full exploitation of metric-free geodesics will Track 2 admit the Einstein metric back into the scene (Chapters 13–15).

But to forego use of the metric is a luxury which Track 1 can ill afford; too little time would be left for relativistic stars, cosmology, black holes, gravitational waves, experimental tests, and the dynamics of geometry. Therefore, the Track-1 overview in this chapter keeps the Einstein metric throughout. But in doing so, it pays a heavy price: (1) no possibility of seeing curvature at work in Newtonian spacetime (Chapter 12); (2) no possibility of comparing and contrasting the geometric structures of Newtonian spacetime (Chapter 12) and Einstein spacetime (Chapter 13), and hence no possibility of grasping fully the Newtonian-based motivation for the Einstein field equations (Chapter 17); (3) no possibility of understanding *fully* the mathematical interrelationships of "geodesic," "parallel transport," "covariant derivative," "curvature," and "metric" (Chapters 9, 10, 11, 13); (4) no possibility of introducing the mathematical subjects *"differential topology"* (geometry without metric or covariant

What the Track-1 reader will miss

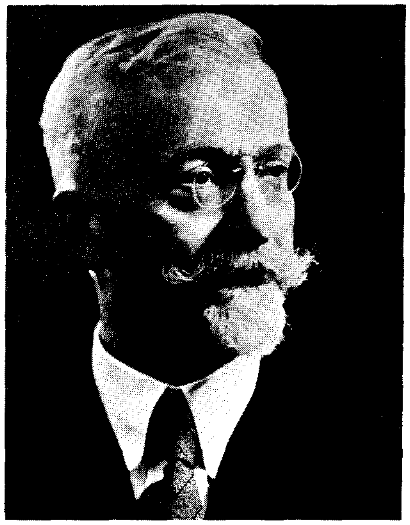
derivative, Chapter 9) and *"affine geometry"* (geometry with covariant derivative but no metric, Chapters 10 and 11), subjects which find major application in modern analytical mechanics [see, e.g., Arnold and Avez (1968); also exercise 4.11 of this book], in Lie group theory with its deep implications for elementary particle physics [see, e.g., Hermann (1966); also exercises 9.12, 9.13, 10.16, and 11.12 of this book], in the theory and solution of partial differential equations [see, e.g., Sternberg (1969)], and, of course, in gravitation theory.

§8.3. THREE ASPECTS OF GEOMETRY: PICTORIAL, ABSTRACT, COMPONENT

Geometry from three viewpoints: pictorial, abstract, component

Gain the power in §8.4 and Chapter 9 to discuss tangent vectors, 1-forms, tensors in curved spacetime; gain the power in §8.5 and Chapter 10 to parallel-transport vectors, to differentiate them, to discuss geodesics; use this power in §8.7 and Chapter 11 to discuss geodesic deviation, to define curvature; .... But full power this will be only if it can be exercised in three ways: in pictures, in abstract notation, and in component notation (Box 8.3). Élie Cartan (Box 8.2) gave new insight into both

Box 8.2 ÉLIE CARTAN, 1869–1951



Élie Cartan is a most remarkable figure in recent mathematical history. One learns from his obituary [Chern and Chevalley (1952)] that he was born a blacksmith's son in southern France, and proved the value of government scholarship aid by rising through the system to a professorship at the Sorbonne in 1912 when he was 43. At the age of 32 he invented the exterior derivative [Cartan (1901)], which he used then mostly in differential equations and the theory of Lie groups, where he had already made significant contributions. He was about fifty when he began applying it to geometry, and sixty before Riemannian geometry specifically was the object of this research, including his text [Cartan (1928)], which is still reprinted and worth studying. Although universally recognized, his work did not find responsive readers until he neared retirement around 1940, when the "Bourbaki" generation of French mathematicians began to provide a conceptual framework for (among other things) Cartan's insights and methods. This made Cartan communicable and teachable as his own writings never were, so that by the time of his death at 82 in 1951 his influence was obviously dominating the revolutions then in full swing in all the fields (Lie groups, differential equations, and differential geometry) in which he had primarily worked.

The modern, abstract, coordinate-free approach to geometry, which is used extensively in this book, is due largely to Élie Cartan. He also discovered the geometric approach to Newtonian gravity that is developed and exploited in Chapter 12.

Box 8.3 THREE LEVELS OF DIFFERENTIAL GEOMETRY

(1) Purely *pictorial* treatment of geometry:
tangent vector is conceived in terms of the separation of two points in the limit in which the points are indefinitely close;
vectors are added and subtracted locally as in flat space;
vectors at distinct points are compared by parallel transport from one point to another;
this parallel transport is accomplished by a "Schild's ladder construction" of geodesics (Box 10.2);
diagrams, yes; algebra, no;
it is tied conceptually as closely as possible to the world of test particles and measurements.

(2) *Abstract* differential geometry:
treats a tangent vector as existing in its own right, without necessity to give its breakdown into components,

$$\boldsymbol{A} = A^0\boldsymbol{e}_0 + A^1\boldsymbol{e}_1 + A^2\boldsymbol{e}_2 + A^3\boldsymbol{e}_3,$$

just as one is accustomed nowadays in electromagnetism to treat the electric vector $\boldsymbol{E}$, without having to write out its components;
uses a similar approach to differentiation (compare gradient operator $\boldsymbol{\nabla}$ of elementary vector analysis, as distinguished from coordinate-dependent pieces of such an operator, such as $\partial/\partial x$, $\partial/\partial y$, etc.);
is the quickest, simplest mathematical scheme one knows to derive general results in differential geometry.

(3) Differential geometry as expressed in the language of *components:*
is indispensible in programming large parts of general relativity for a computer;
is convenient or necessary or both when one is dealing even at the level of elementary algebra with the most simple applications of relativity, from the expansion of the Friedmann universe to the curvature around a static center of attraction.

Newtonian gravity (Chapter 12) and the central geometric simplicity of Einstein's field equations (Chapter 15), because he had full command of all three methods of doing differential geometry. Today, no one has full power to communicate with others about the subject who cannot express himself in all three languages. Hence the interplay between the three forms of expression in what follows.

Planetary orbit as example of three viewpoints

It is not new to go back and forth between the three languages, as witnesses the textbook treatment of the velocity and acceleration of a planet in Kepler motion around the sun. The velocity is written

$$\boldsymbol{v} = v^{\hat{r}}\boldsymbol{e}_{\hat{r}} + v^{\hat{\phi}}\boldsymbol{e}_{\hat{\phi}}. \tag{8.1}$$

(The hats "^" on $\boldsymbol{e}_{\hat{r}}$ and $\boldsymbol{e}_{\hat{\phi}}$ signify that these are unit vectors.) The acceleration is

$$\boldsymbol{a} = \frac{d\boldsymbol{v}}{dt} = \frac{dv^{\hat{r}}}{dt}\boldsymbol{e}_{\hat{r}} + \frac{dv^{\hat{\phi}}}{dt}\boldsymbol{e}_{\hat{\phi}} + v^{\hat{r}}\frac{d\boldsymbol{e}_{\hat{r}}}{dt} + v^{\hat{\phi}}\frac{d\boldsymbol{e}_{\hat{\phi}}}{dt}. \tag{8.2}$$

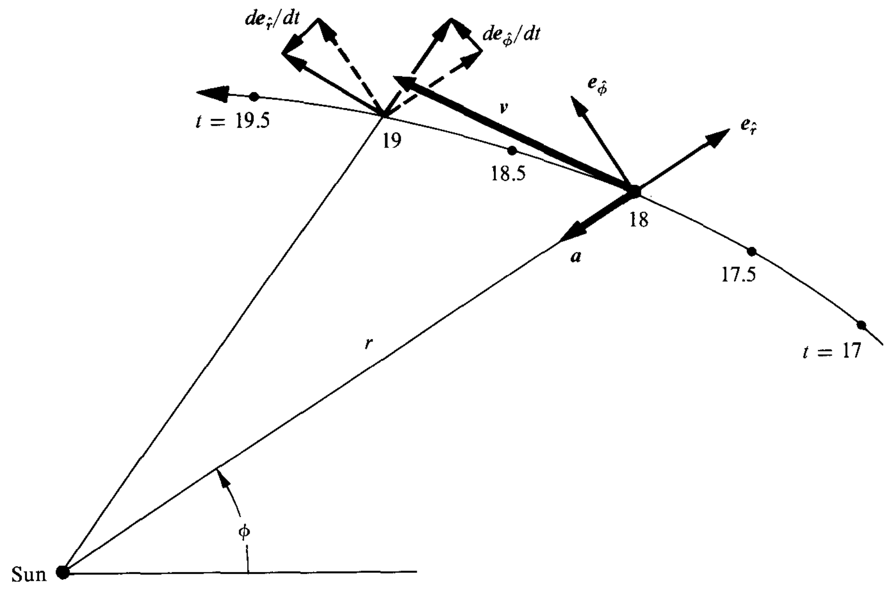


Figure 8.1.
A Keplerian orbit in the sun's gravitational field, as treated using the standard Euclidean-space version of Newtonian gravity. The basis vectors themselves change from point to point along the orbit [equations (8.3)]. This figure illustrates the pictorial aspect of differential geometry. Later (exercise 8.5) it will illustrate the concepts of "covariant derivative" and "connection coefficients."

The unit vectors are turning (Figure 8.1) with the angular velocity $\omega = d\phi/dt$; so

$$\begin{aligned} \frac{de_{\hat{r}}}{dt} &= \omega e_{\hat{\phi}} = \frac{d\phi}{dt} e_{\hat{\phi}}, \\ \frac{de_{\hat{\phi}}}{dt} &= -\omega e_{\hat{r}} = -\frac{d\phi}{dt} e_{\hat{r}}. \end{aligned} \tag{8.3}$$

Thus the components of the acceleration have the values

$$a^{\hat{r}} = \frac{dv^{\hat{r}}}{dt} - v^{\hat{\phi}} \frac{d\phi}{dt} = \frac{d^2 r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 \tag{8.4a}$$

and

$$a^{\hat{\phi}} = \frac{dv^{\hat{\phi}}}{dt} + v^{\hat{r}} \frac{d\phi}{dt} = \frac{d}{dt}\left(r \frac{d\phi}{dt}\right) + \frac{dr}{dt}\frac{d\phi}{dt}. \tag{8.4b}$$

Here is the acceleration in the language of components; $\boldsymbol{a}$ was the acceleration in abstract language; and Figure 8.1 shows the acceleration as an arrow. Each of these three languages will receive its natural generalization in the coming sections and chapters from two-dimensional flat space (with curvilinear coordinates) to four-dimensional curved spacetime, and from spacetime to more general manifolds (see §9.7 on manifolds).

Turn now to the Track-1 overview of differential geometry.

§8.4. TENSOR ALGEBRA IN CURVED SPACETIME

To see spacetime curvature at work, examine tidal gravitational forces (geodesic deviation); and to probe these forces, make measurements in a finite-sized laboratory. Squeeze the laboratory to infinitesimal size; all effects of spacetime curvature become infinitesimal; the physicist cannot tell whether he is in flat spacetime or curved spacetime. Neither can the mathematician, in the limit as his domain of attention squeezes down to a single event, $\mathscr{P}_0$.

Tensor algebra:

(1) occurs in infinitesimal neighborhood of an event

At the event $\mathscr{P}_0$ (in the infinitesimal laboratory) both physicist and mathematician can talk of vectors, of 1-forms, of tensors; and no amount of spacetime curvature can force the discussion to change from its flat-space form. A particle at $\mathscr{P}_0$ has a 4-momentum $\boldsymbol{p}$, with squared length

(2) is same in curved spacetime as in flat

$$\boldsymbol{p}^2 = \boldsymbol{p}\cdot\boldsymbol{p} = \boldsymbol{g}(\boldsymbol{p},\boldsymbol{p}) = -m^2.$$

The squared length, as always, is calculated by inserting $\boldsymbol{p}$ into both slots of a linear machine, the metric $\boldsymbol{g}$ at $\mathscr{P}_0$. The particle also has a 4-acceleration $\boldsymbol{a}$ at $\mathscr{P}_0$; and, if the particle is charged and freely moving, then $\boldsymbol{a}$ is produced by the electromagnetic field tensor $\boldsymbol{F}$:

$$m\boldsymbol{a} = e\boldsymbol{F}(\ldots,\boldsymbol{u}).$$

In no way can curvature affect such local, coordinate-free, geometric relations. And in no way can it prevent one from introducing a local Lorentz frame at $\mathscr{P}_0$, and from performing standard, flat-space index manipulations in it:

$$\boldsymbol{p} = p^\alpha \boldsymbol{e}_\alpha, \qquad \boldsymbol{p}^2 = p^\alpha p^\beta \eta_{\alpha\beta} = p^\alpha p_\alpha, \qquad m a^\alpha = eF^{\alpha\beta}u_\beta.$$

But local Lorentz frames are not enough for the man who would calculate in curved spacetime. Non-Lorentz frames (nonorthonormal basis vectors $\{\boldsymbol{e}_\alpha\}$) often simplify calculations. Fortunately, no effort at all is required to master the rules of "index mechanics" in an arbitrary basis at a fixed event $\mathscr{P}_0$. The rules are identical to those in flat spacetime, except that (1) the covariant Lorentz components $\eta_{\alpha\beta}$ of the metric are replaced by

(3) rules for component manipulation change slightly when using nonorthonormal basis

Components of metric

$$g_{\alpha\beta} \equiv \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta \equiv \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta); \tag{8.5}$$

(2) the contravariant components $\eta^{\alpha\beta}$ are replaced by $g^{\alpha\beta}$, where

$$\|g^{\alpha\beta}\| \equiv \|g_{\alpha\beta}\|^{-1} \text{ (matrix inverse)}; \tag{8.6}$$

i.e.,

$$g_{\alpha\beta}g^{\beta\gamma} = \delta_\alpha{}^\gamma; \tag{8.6$'$}$$

(3) the Lorentz transformation matrix $\|\Lambda^{\alpha'}{}_\beta\|$ and its inverse $\|\Lambda^\beta{}_{\alpha'}\|$ are replaced by an arbitrary but nonsingular transformation matrix $\|L^{\alpha'}{}_\beta\|$ and its inverse $\|L^\beta{}_{\alpha'}\|$:

Transformation of basis

$$\boldsymbol{e}_\beta = \boldsymbol{e}_{\alpha'} L^{\alpha'}{}_\beta, \qquad p^\beta = L^\beta{}_{\alpha'} p^{\alpha'}, \tag{8.7}$$

$$\|L^\beta{}_{\alpha'}\| = \|L^{\alpha'}{}_\beta\|^{-1}; \tag{8.8}$$

(4) in the special case of "coordinate bases," $\boldsymbol{e}_\alpha = \partial\mathscr{P}/\partial x^\alpha$, $\boldsymbol{e}_{\beta'} = \partial\mathscr{P}/\partial x^{\beta'}$,

$$L^{\alpha'}{}_{\beta} = \partial x^{\alpha'}/\partial x^{\beta}, \qquad L^{\beta}{}_{\alpha'} = \partial x^{\beta}/\partial x^{\alpha'}; \tag{8.9}$$

Components of Levi-Civita tensor

and (5) the Levi-Civita tensor $\boldsymbol{\varepsilon}$, like the metric tensor, has components that depend on how nonorthonormal the basis vectors are (see exercise 8.3): if $\boldsymbol{e}_0$ points toward the future and $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ are righthanded, then

$$\begin{aligned} \varepsilon_{\alpha\beta\gamma\delta} &= (-g)^{1/2}[\alpha\beta\gamma\delta], \\ \varepsilon^{\alpha\beta\gamma\delta} &= g^{-1}\varepsilon_{\alpha\beta\gamma\delta} = -(-g)^{-1/2}[\alpha\beta\gamma\delta], \end{aligned} \tag{8.10a}$$

where $[\alpha\beta\gamma\delta]$ is the completely antisymmetric symbol

$$[\alpha\beta\gamma\delta] \equiv \begin{cases} +1 \text{ if } \alpha\beta\gamma\delta \text{ is an even permutation of } 0123, \\ -1 \text{ if } \alpha\beta\gamma\delta \text{ is an odd permutation of } 0123, \\ \;\;\,0 \text{ if } \alpha\beta\gamma\delta \text{ are not all different,} \end{cases} \tag{8.10b}$$

and where g is the determinant of the matrix $\|g_{\alpha\beta}\|$

$$g \equiv \det\|g_{\alpha\beta}\| = \det\|\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta\|. \tag{8.11}$$

Read Box 8.4 for full discussion and proofs; work exercise 8.1 below for fuller understanding and mastery.

Several dangers are glossed over in this discussion. In flat spacetime one often does not bother to say where a vector, 1-form, or tensor is located. One freely moves geometric objects from event to event without even thinking. Of course, the unwritten rule of transport is: hold all lengths and directions fixed while moving; i.e., hold all Lorentz-frame components fixed; i.e., "parallel-transport" the object. But in

Box 8.4 TENSOR ALGEBRA AT A FIXED EVENT IN AN *ARBITRARY* BASIS

A. Bases

Tangent-vector basis: Pick $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ at $\mathscr{P}_0$ arbitrarily—but insist they be linearly independent.

"Dual basis" for 1-forms: The basis $\{\boldsymbol{e}_\alpha\}$ determines a 1-form basis $\{\boldsymbol{\omega}^\alpha\}$ (its "dual basis") by

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta$$

[see equation (2.19)].

Geometric interpretation (Figure 9.2): $\boldsymbol{e}_2$, $\boldsymbol{e}_3$, $\boldsymbol{e}_0$ lie parallel to surfaces of $\boldsymbol{\omega}^1$; $\boldsymbol{e}_1$ pierces precisely one surface of $\boldsymbol{\omega}^1$.

Function interpretation: $\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta$ determines the value of $\boldsymbol{\omega}^\alpha$ on any vector $\boldsymbol{u} = u^\beta \boldsymbol{e}_\beta$ (number of "bongs of bell" as $\boldsymbol{u}$ pierces $\boldsymbol{\omega}^\alpha$):

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{u}\rangle = \langle \boldsymbol{\omega}^\alpha, u^\beta \boldsymbol{e}_\beta\rangle = u^\beta \langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta\rangle = u^\beta \delta^\alpha{}_\beta = u^\alpha.$$

Special case: *coordinate bases.* Choose an arbitrary coordinate system $\{x^\alpha(\mathscr{P})\}$. At $\mathscr{P}_0$ choose $\boldsymbol{e}_\alpha = \partial\mathscr{P}/\partial x^\alpha$ as basis vectors. Then the dual basis is $\boldsymbol{\omega}^\alpha = \boldsymbol{d}x^\alpha$. *Proof:* the general coordinate-free relation $\langle \boldsymbol{d}f, \boldsymbol{v}\rangle = \partial_{\boldsymbol{v}} f$ [equation (2.17)], with $f = x^\alpha$ and $\boldsymbol{v} = \partial\mathscr{P}/\partial x^\beta$, reads

$$\langle \boldsymbol{d}x^\alpha, \partial\mathscr{P}/\partial x^\beta\rangle = (\partial/\partial x^\beta)x^\alpha = \delta^\alpha{}_\beta.$$

B. Algebra of Tangent Vectors and 1-Forms

The Lorentz-frame discussion of equations (2.19) to (2.22) is completely unchanged when one switches to an arbitrary basis. Its conclusions:

expansion, $\boldsymbol{u} = \boldsymbol{e}_\alpha u^\alpha$, $\boldsymbol{\sigma} = \sigma_\alpha \boldsymbol{\omega}^\alpha$;

calculation of components, $u^\alpha = \langle \boldsymbol{\omega}^\alpha, \boldsymbol{u}\rangle$, $\sigma_\alpha = \langle \boldsymbol{\sigma}, \boldsymbol{e}_\alpha\rangle$;

value of form on vector, $\langle \boldsymbol{\sigma}, \boldsymbol{u}\rangle = \sigma_\alpha u^\alpha$.

Application to gradients of functions:

expansion, $\boldsymbol{d}f = f_{,\alpha}\boldsymbol{\omega}^\alpha$ [defines $f_{,\alpha}$];

calculation of components, $f_{,\alpha} = \langle \boldsymbol{d}f, \boldsymbol{e}_\alpha\rangle = \partial_{\boldsymbol{e}_\alpha} f$ [see equation (2.17)].

Raising and lowering of indices is accomplished with $g_{\alpha\beta}$ and $g^{\alpha\beta}$ [equations (8.5) and (8.6)]. Proof:

$\tilde{\boldsymbol{u}}$, the 1-form corresponding to $\boldsymbol{u}$, is defined by $\langle \tilde{\boldsymbol{u}}, \boldsymbol{v}\rangle = \boldsymbol{u}\cdot\boldsymbol{v}$ for all $\boldsymbol{v}$;

thus, $u_\alpha \equiv \langle \tilde{\boldsymbol{u}}, \boldsymbol{e}_\alpha\rangle = \boldsymbol{u}\cdot\boldsymbol{e}_\alpha = u^\beta \boldsymbol{e}_\beta \cdot \boldsymbol{e}_\alpha = u^\beta g_{\beta\alpha}$;

inverting this equation yields $u^\beta = g^{\beta\alpha}u_\alpha$.

C. Change of Basis

The discussion of Lorentz transformations in equations (2.39) to (2.43) is applicable to general changes of basis if one replaces $\|\Lambda^{\alpha'}{}_\beta\|$ by an arbitrary but nonsingular matrix $\|L^{\alpha'}{}_\beta\|$ [equations (8.7), (8.8)]. Conclusions:

$$\begin{aligned}
\boldsymbol{e}_{\alpha'} &= \boldsymbol{e}_\beta L^\beta{}_{\alpha'}, & \boldsymbol{e}_\beta &= \boldsymbol{e}_{\alpha'} L^{\alpha'}{}_\beta;\\
\boldsymbol{\omega}^{\alpha'} &= L^{\alpha'}{}_\beta \boldsymbol{\omega}^\beta, & \boldsymbol{\omega}^\beta &= L^\beta{}_{\alpha'}\boldsymbol{\omega}^{\alpha'};\\
v^{\alpha'} &= L^{\alpha'}{}_\beta v^\beta, & v^\beta &= L^\beta{}_{\alpha'} v^{\alpha'};\\
\sigma_{\alpha'} &= \sigma_\beta L^\beta{}_{\alpha'}, & \sigma_\beta &= \sigma_{\alpha'} L^{\alpha'}{}_\beta.
\end{aligned}$$

When both bases are coordinate bases, then $L^\beta{}_{\alpha'} = \partial x^\beta/\partial x^{\alpha'}$, $L^{\alpha'}{}_\beta = \partial x^{\alpha'}/\partial x^\beta$. *Proof:*

$$\boldsymbol{e}_{\alpha'} = \frac{\partial}{\partial x^{\alpha'}} = \frac{\partial x^\beta}{\partial x^{\alpha'}}\frac{\partial}{\partial x^\beta} = \frac{\partial x^\beta}{\partial x^{\alpha'}}\boldsymbol{e}_\beta; \qquad \text{similarly } \boldsymbol{e}_\beta = \frac{\partial x^{\alpha'}}{\partial x^\beta}\boldsymbol{e}_{\alpha'}.$$

Box 8.4 (*continued*)

D. Algebra of Tensors

The discussions of tensor algebra given in §3.2 [equations (3.8) to (3.22)] and in §3.5 (excluding gradient and divergence) are unchanged, except that

$$\eta_{\alpha\beta} \longrightarrow g_{\alpha\beta} \qquad \eta^{\alpha\beta} \longrightarrow g^{\alpha\beta}, \qquad \Lambda^{\alpha'}{}_{\beta} \longrightarrow L^{\alpha'}{}_{\beta}, \qquad \Lambda^{\beta}{}_{\alpha'} \longrightarrow L^{\beta}{}_{\alpha'};$$

and the components of the Levi-Civita tensor are changed from (3.50) to (8.10) [see exercise 8.3[.

Chief conclusions:

expansion, $\boldsymbol{S} = S^{\alpha}{}_{\beta\gamma}\boldsymbol{e}_{\alpha} \otimes \boldsymbol{\omega}^{\beta} \otimes \boldsymbol{\omega}^{\gamma}$;
components, $S^{\alpha}{}_{\beta\gamma} = \boldsymbol{S}(\boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta}, \boldsymbol{e}_{\gamma})$;
raising and lowering indices, $S_{\mu\beta}{}^{\nu} = g_{\mu\alpha}g^{\nu\gamma}S^{\alpha}{}_{\beta\gamma}$;
change of basis, $S^{\lambda'}{}_{\mu'\nu'} = L^{\lambda'}{}_{\alpha}L^{\beta}{}_{\mu'}L^{\gamma}{}_{\nu'}S^{\alpha}{}_{\beta\gamma}$;
machine operation, $\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{u}, \boldsymbol{v}) = S^{\alpha}{}_{\beta\gamma}\sigma_{\alpha}u^{\beta}v^{\gamma}$;
tensor product, $\boldsymbol{T} = \boldsymbol{u} \otimes \boldsymbol{v} \Longleftrightarrow T^{\alpha\beta} = u^{\alpha}v^{\beta}$;
contraction, "$\boldsymbol{M}$ = contraction of $\boldsymbol{R}$ on slots 1 and 3" $\Longleftrightarrow M_{\mu\nu} = R^{\alpha}{}_{\mu\alpha\nu}$;
wedge product, $\boldsymbol{\alpha} \wedge \boldsymbol{\beta}$ has components $\alpha^{\mu}\beta^{\nu} - \beta^{\mu}\alpha^{\nu}$;
Dual, ${}^*J_{\alpha\beta\gamma} = J^{\mu}\varepsilon_{\mu\alpha\beta\gamma}$, ${}^*F_{\alpha\beta} = \frac{1}{2}F^{\mu\nu}\varepsilon_{\mu\nu\alpha\beta}$, ${}^*B_{\alpha} = \frac{1}{6}B^{\lambda\mu\nu}\varepsilon_{\lambda\mu\nu\alpha}$.

E. Commutators (exercise 8.2; §9.6; Box 9.2)

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are tangent vector fields, one takes the view that $\boldsymbol{u} = \partial_{\boldsymbol{u}}$ and $\boldsymbol{v} = \partial_{\boldsymbol{v}}$, and one defines

$$[\boldsymbol{u}, \boldsymbol{v}] \equiv [\partial_{\boldsymbol{u}}, \partial_{\boldsymbol{v}}] \equiv \partial_{\boldsymbol{u}}\,\partial_{\boldsymbol{v}} - \partial_{\boldsymbol{v}}\,\partial_{\boldsymbol{u}}.$$

This commutator is itself a tangent vector field.

Components in a coordinate basis:

$$[\boldsymbol{u}, \boldsymbol{v}] = (u^{\beta}v^{\alpha}{}_{,\beta} - v^{\beta}u^{\alpha}{}_{,\beta})\underbrace{(\partial/\partial x^{\alpha})}_{[=\boldsymbol{e}_{\alpha}]}.$$

Commutation coefficients of a basis:

$$[\boldsymbol{e}_{\alpha}, \boldsymbol{e}_{\beta}] \equiv c_{\alpha\beta}{}^{\gamma}\boldsymbol{e}_{\gamma}, \qquad c_{\alpha\beta\mu} \equiv c_{\alpha\beta}{}^{\gamma}g_{\gamma\mu}.$$

Coordinate basis ("holonomic") $c_{\alpha\beta}{}^{\gamma} = 0$;
Noncoordinate basis ("anholonomic") some $c_{\alpha\beta}{}^{\gamma} \neq 0$ (see exercise 9.9).

curved spacetime there is no global Lorentz coordinate system in which to hold components fixed; and objects initially parallel, after "parallel transport" along different curves cease to be parallel ("geodesic deviation"; Earth's meridians, parallel at equator, cross at north and south poles). Thus, in curved spacetime one must not blithely move a geometric object from point to point, without carefully specifying how it is to be moved and by what route. Each local geometric object has its own official place of residence (event $\mathscr{P}_0$); it can interact with other objects residing there (tensor algebra); but it cannot interact with any object at another event $\mathscr{Q}$, until it has been carefully transported from $\mathscr{P}_0$ to $\mathscr{Q}$.

Vectors and tensors must not be moved blithely from point to point

This line of reasoning, pursued further, leads one to speak of the *"tangent space"* at each event, in which that event's vectors (arrows) and 1-forms (families of surfaces) lie, and in which its tensors (linear machines) operate. One even draws heuristic pictures of the tangent space, as in Figure 9.1 (p. 231).

Tangent space defined

Another danger in curved spacetime is the temptation to regard vectors as arrows linking two events ("point for head and point for tail")—i.e., to regard the tangent space of Figure 9.1 as lying in spacetime itself. This practice can be useful for heuristic purposes, but it is incompatible with complete mathematical precision. (How is the tangent space to be molded into a warped surface?) Four definitions of a vector were given in Figure 2.1 (page 49): three definitions relying on "point for head and point for tail"; one, "$d\mathscr{P}/d\lambda$", purely local. Only the local definition is wholly viable in curved spacetime, and even it can be improved upon, in the eyes of mathematicians, as follows.

Definitions of vector in curved spacetime:

(1) as $d\mathscr{P}/d\lambda$

There is a one-to-one correspondence (complete "isomorphism") between vectors $\boldsymbol{u}$ and directional derivative operators $\partial_{\boldsymbol{u}}$. The concept of vector is a bit fuzzy, but "directional derivative" is perfectly well-defined. To get rid of all fuzziness, exploit the isomorphism to the full: *define* the tangent vector $\boldsymbol{u}$ to be equal to the corresponding directional derivative

(2) as directional derivative

$$\boldsymbol{u} \equiv \partial_{\boldsymbol{u}}. \tag{8.12}$$

(This practice, unfamiliar as it may be to a physicist at first, has mathematical power; so this book will use it frequently. For a fuller discussion, see §9.2.)

EXERCISES

Exercise 8.1. PRACTICE WITH TENSOR ALGEBRA

Let t, x, y, z be Lorentz coordinates in flat spacetime, and let

$$r = (x^2 + y^2 + z^2)^{1/2}, \qquad \theta = \cos^{-1}(z/r), \qquad \phi = \tan^{-1}(y/x)$$

be the corresponding spherical coordinates. Then

$$\boldsymbol{e}_0 = \partial\mathscr{P}/\partial t, \qquad \boldsymbol{e}_r = \partial\mathscr{P}/\partial r, \qquad \boldsymbol{e}_\theta = \partial\mathscr{P}/\partial\theta, \qquad \boldsymbol{e}_\phi = \partial\mathscr{P}/\partial\phi$$

is a coordinate basis, and

$$\boldsymbol{e}_{\hat{0}} = \frac{\partial\mathscr{P}}{\partial t}, \qquad \boldsymbol{e}_{\hat{r}} = \frac{\partial\mathscr{P}}{\partial r}, \qquad \boldsymbol{e}_{\hat{\theta}} = \frac{1}{r}\frac{\partial\mathscr{P}}{\partial\theta}, \qquad \boldsymbol{e}_{\hat{\phi}} = \frac{1}{r\sin\theta}\frac{\partial\mathscr{P}}{\partial\phi}$$

is a noncoordinate basis.

(a) Draw a picture of $\boldsymbol{e}_\theta$, $\boldsymbol{e}_\phi$, $\boldsymbol{e}_{\hat\theta}$, and $\boldsymbol{e}_{\hat\phi}$ at several different points on a sphere of constant t, r. [Answer for $\boldsymbol{e}_\theta$, $\boldsymbol{e}_\phi$ should resemble Figure 9.1.]

(b) What are the 1-form bases $\{\boldsymbol{\omega}^\alpha\}$ and $\{\boldsymbol{\omega}^{\hat\alpha}\}$ dual to these tangent-vector bases? [*Answer:* $\boldsymbol{\omega}^0 = \boldsymbol{d}t$, $\boldsymbol{\omega}^r = \boldsymbol{d}r$, $\boldsymbol{\omega}^\theta = \boldsymbol{d}\theta$, $\boldsymbol{\omega}^\phi = \boldsymbol{d}\phi$; $\boldsymbol{\omega}^{\hat 0} = \boldsymbol{d}t$, $\boldsymbol{\omega}^{\hat r} = \boldsymbol{d}r$, $\boldsymbol{\omega}^{\hat\theta} = r\,\boldsymbol{d}\theta$, $\boldsymbol{\omega}^{\hat\phi} = r \sin\theta\, \boldsymbol{d}\phi$.]

(c) What is the transformation matrix linking the original Lorentz frame to the spherical coordinate frame $\{\boldsymbol{e}_\alpha\}$? [*Answer:* nonzero components are

$$L^t{}_0 = 1, \qquad L^z{}_r = \frac{\partial z}{\partial r} = \cos\theta, \qquad L^z{}_\theta = \frac{\partial z}{\partial \theta} = -r\sin\theta,$$

$$L^x{}_r = \sin\theta\cos\phi, \qquad L^x{}_\theta = r\cos\theta\cos\phi, \qquad L^x{}_\phi = -r\sin\theta\sin\phi,$$

$$L^y{}_r = \sin\theta\sin\phi, \qquad L^y{}_\theta = r\cos\theta\sin\phi, \qquad L^y{}_\phi = r\sin\theta\cos\phi.]$$

(d) Use this transformation matrix to calculate the metric components $g_{\alpha\beta}$ in the spherical coordinate basis, and invert the resulting matrix to get $g^{\alpha\beta}$. [*Answer:*

$$g_{00} = -1, \quad g_{rr} = 1, \quad g_{\theta\theta} = r^2, \quad g_{\phi\phi} = r^2\sin^2\theta, \quad \text{all other } g_{\alpha\beta} = 0.$$
$$g^{00} = -1, \quad g^{rr} = 1, \quad g^{\theta\theta} = r^{-2}, \quad g^{\phi\phi} = r^{-2}\sin^{-2}\theta, \quad \text{all other } g^{\alpha\beta} = 0.]$$

(e) Show that the noncoordinate basis $\{\boldsymbol{e}_{\hat\alpha}\}$ is orthonormal everywhere; i.e., that $g_{\hat\alpha\hat\beta} = \eta_{\alpha\beta}$; i.e. that

$$\boldsymbol{g} = -\boldsymbol{\omega}^{\hat 0}\otimes\boldsymbol{\omega}^{\hat 0} + \boldsymbol{\omega}^{\hat r}\otimes\boldsymbol{\omega}^{\hat r} + \boldsymbol{\omega}^{\hat\theta}\otimes\boldsymbol{\omega}^{\hat\theta} + \boldsymbol{\omega}^{\hat\phi}\otimes\boldsymbol{\omega}^{\hat\phi}.$$

(f) Write the gradient of a function f in terms of the spherical coordinate and noncoordinate bases. [*Answer:*

$$\begin{aligned}\boldsymbol{d}f &= \frac{\partial f}{\partial t}\boldsymbol{d}t + \frac{\partial f}{\partial r}\boldsymbol{d}r + \frac{\partial f}{\partial \theta}\boldsymbol{d}\theta + \frac{\partial f}{\partial \phi}\boldsymbol{d}\phi \\ &= \frac{\partial f}{\partial t}\boldsymbol{\omega}^{\hat 0} + \frac{\partial f}{\partial r}\boldsymbol{\omega}^{\hat r} + \frac{1}{r}\frac{\partial f}{\partial \theta}\boldsymbol{\omega}^{\hat\theta} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\boldsymbol{\omega}^{\hat\phi}.]\end{aligned}$$

(g) What are the components of the Levi-Civita tensor in the spherical coordinate and noncoordinate bases? [*Answer for coordinate basis:*

$$\varepsilon_{0r\theta\phi} = -\varepsilon_{r0\theta\phi} = +\varepsilon_{r\theta0\phi} = \cdots = r^2\sin\theta,$$
$$\varepsilon^{0r\theta\phi} = -\varepsilon^{r0\theta\phi} = +\varepsilon^{r\theta\phi0} = \cdots = -r^{-2}\sin^{-1}\theta.]$$

Exercise 8.2. COMMUTATORS

Take the mathematician's viewpoint that tangent vectors and directional derivatives are the same thing, $\boldsymbol{u} \equiv \partial_{\boldsymbol{u}}$. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vector fields, and define their commutator in the manner familiar from quantum mechanics

$$[\boldsymbol{u}, \boldsymbol{v}] \equiv [\partial_{\boldsymbol{u}}, \partial_{\boldsymbol{v}}] \equiv \partial_{\boldsymbol{u}}\,\partial_{\boldsymbol{v}} - \partial_{\boldsymbol{v}}\,\partial_{\boldsymbol{u}}. \tag{8.13a}$$

(a) Derive the following expression for $[\boldsymbol{u}, \boldsymbol{v}]$, valid in any *coordinate* basis,

$$[\boldsymbol{u}, \boldsymbol{v}] = (u^\beta v^\alpha{}_{,\beta} - v^\beta u^\alpha{}_{,\beta})\boldsymbol{e}_\alpha. \tag{8.13b}$$

Thus, despite the fact that it looks like a second-order differential operator, $[\boldsymbol{u}, \boldsymbol{v}]$ is actually of first order—i.e., it is a tangent vector.

(b) For any basis $\{\boldsymbol{e}_\alpha\}$, one defines the "*commutation coefficients*" $c_{\beta\gamma}{}^\alpha$ and $c_{\beta\gamma\alpha}$ by

$$[\boldsymbol{e}_\beta, \boldsymbol{e}_\gamma] \equiv c_{\beta\gamma}{}^\alpha\boldsymbol{e}_\alpha; \qquad c_{\beta\gamma\alpha} \equiv g_{\alpha\mu}c_{\beta\gamma}{}^\mu. \tag{8.14}$$

Show that $c_{\beta\gamma}{}^\alpha = c_{\beta\gamma\alpha} = 0$ for any coordinate basis.

(c) Calculate $c_{\hat\beta\hat\gamma}{}^{\hat\alpha}$ for the spherical noncoordinate basis of exercise 8.1. [*Answer:* All vanish except

$$c_{\hat r\hat\theta}{}^{\hat\theta} = -c_{\hat\theta\hat r}{}^{\hat\theta} = -1/r,$$
$$c_{\hat r\hat\phi}{}^{\hat\phi} = -c_{\hat\phi\hat r}{}^{\hat\phi} = -1/r,$$
$$c_{\hat\theta\hat\phi}{}^{\hat\phi} = -c_{\hat\phi\hat\theta}{}^{\hat\phi} = -\cot\theta/r.]$$

Exercise 8.3. COMPONENTS OF LEVI-CIVITA TENSOR IN NONORTHONORMAL FRAME

(a) Show that expressions (8.10) are the components of $\boldsymbol{\varepsilon}$ in an arbitrary basis, with $\boldsymbol{e}_0$ pointing toward the future and $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ right-handed. [Hints: (1) Review the discussion of $\boldsymbol{\varepsilon}$ in Lorentz frames, given in exercise 3.13. (2) Calculate $\varepsilon_{\alpha\beta\gamma\delta}$ and $\varepsilon^{\alpha\beta\gamma\delta}$ by transforming from a local Lorentz frame $\{\boldsymbol{e}_{\hat\mu}\}$, e.g.,

$$\varepsilon_{\alpha\beta\gamma\delta} = L^{\hat\mu}{}_{\alpha}L^{\hat\nu}{}_{\beta}L^{\hat\lambda}{}_{\gamma}L^{\hat\rho}{}_{\delta}\varepsilon_{\hat\mu\hat\nu\hat\lambda\hat\rho}.$$

(3) Show that these expressions reduce to

$$\varepsilon_{\alpha\beta\gamma\delta} = \det\|L^{\hat\mu}{}_{\nu}\|\varepsilon_{\hat\alpha\hat\beta\hat\gamma\hat\delta}, \qquad \varepsilon^{\alpha\beta\gamma\delta} = \det\|L^{\nu}{}_{\hat\mu}\|\varepsilon^{\hat\alpha\hat\beta\hat\gamma\hat\delta}.$$

(4) Show, from the transformation law for the metric components, that

$$(\det\|L^{\nu}{}_{\hat\mu}\|)^2 \det\|g_{\alpha\beta}\| = -1.$$

(5) Combine these results to obtain expressions (8.10).]

(b) Show that the components of the permutation tensors [defined by equations (3.50h)–(3.50j)] have the same values [equations (3.50k)–(3.50m)] in arbitrary frames as in Lorentz frames.

Additional exercises on tensor algebra: exercises 9.3 and 9.4 (page 234).

§8.5. PARALLEL TRANSPORT, COVARIANT DERIVATIVE, CONNECTION COEFFICIENTS, GEODESICS

Equivalence principle as vehicle between flat spacetime and curved

The vehicle that carries one from classical mechanics to quantum mechanics is the correspondence principle. Similarly, the vehicle between flat spacetime and curved spacetime is the equivalence principle: "The laws of physics are the same in any local Lorentz frame of curved spacetime as in a global Lorentz frame of flat spacetime." But to apply the equivalence principle, one must first have a mathematical representation of a local Lorentz frame. The obvious choice is this: *A local Lorentz frame at a given event $\mathcal{P}_0$ is the closest thing there is to a global Lorentz frame at that event; i.e., it is a coordinate system in which*

Local Lorentz frame: mathematical representation

$$g_{\mu\nu}(\mathcal{P}_0) = \eta_{\mu\nu}, \tag{8.15a}$$

and in which $g_{\mu\nu}$ holds as tightly as possible to $\eta_{\mu\nu}$ in the neighborhood of $\mathcal{P}_0$:

$$g_{\mu\nu,\alpha}(\mathcal{P}_0) = 0. \tag{8.15b}$$

More tightly than this it cannot hold in general [$g_{\mu\nu,\alpha\beta}(\mathcal{P}_0)$ cannot be set to zero]; spacetime curvature forces it to change. [Combine §11.5 with equations (8.24) and (8.44).]

Parallel transport defined

An observer in a local Lorentz frame in curved spacetime can compare vectors and tensors at neighboring events, just as he would in flat spacetime. But to make the comparison, he must parallel-transport them to a common event. For him the act of parallel transport is simple: he keeps all Lorentz-frame components fixed, just as if he were in flat spacetime. But for a man without a local Lorentz frame—perhaps with no coordinate system or basis vectors at all—parallel transport is less trivial. He must either ask his Lorentz-based friend the result, or he must use a more sophisticated technique. One technique he can use—a "Schild's ladder" construction that requires no coordinates or basis vectors whatsoever—is described in §10.2 and Box 10.2. But the Track-1 reader need not master Schild's ladder. He can always ask a local Lorentz observer what the result of any given parallel transport is, or he can use general formulas worked out below.

Comparison by parallel transport is the foundation on which rests the gradient of a tensor field, $\boldsymbol{\nabla T}$. No mention of parallel transport was made in §3.5, where the gradient was first defined, but parallel transport occurred implicitly: one defined $\boldsymbol{\nabla T}$ in such a way that its components were $T^{\alpha}{}_{\beta,\gamma} = \partial T^{\alpha}{}_{\beta}/\partial x^{\gamma}$ [for $\boldsymbol{T}$ a $\binom{1}{1}$ tensor]; i.e., one asked $\boldsymbol{\nabla T}$ to measure how much the Lorentz-frame components of $\boldsymbol{T}$ change from point to point. But "no change in Lorentz components" would have meant "parallel transport," so one was implicitly asking for the change in $\boldsymbol{T}$ relative to what $\boldsymbol{T}$ would have been after pure parallel transport.

Covariant derivative defined

To codify in abstract notation this concept of differentiation, proceed as follows. First define the "covariant derivative" $\boldsymbol{\nabla_u T}$ of $\boldsymbol{T}$ along a curve $\mathscr{P}(\lambda)$, whose tangent vector is $\boldsymbol{u} = d\mathscr{P}/d\lambda$:

$$(\boldsymbol{\nabla_u T})_{\text{at }\mathscr{P}(0)} = \lim_{\varepsilon\to 0}\left\{\frac{\boldsymbol{T}[\mathscr{P}(\varepsilon)]_{\text{parallel-transported to }\mathscr{P}(0)} - \boldsymbol{T}[\mathscr{P}(0)]}{\varepsilon}\right\}. \tag{8.16}$$

Gradient defined

(See Figure 8.2 for the special case where $\boldsymbol{T}$ is a vector field $\boldsymbol{v}$.) Then define $\boldsymbol{\nabla T}$ to be the linear machine, that gives $\boldsymbol{\nabla_u T}$ when $\boldsymbol{u}$ is inserted into its last slot:

$$\boldsymbol{\nabla T}(\ldots,\ldots,\boldsymbol{u}) \equiv \boldsymbol{\nabla_u T}. \tag{8.17}$$

The result is the same animal ("gradient") as was defined in §3.5 (for proof see exercise 8.8). But this alternative definition makes clear the relationship to parallel transport, including the fact that

$$\boldsymbol{\nabla_u T} = 0 \iff \boldsymbol{T} \text{ is parallel-transported along } \boldsymbol{u} = d\mathscr{P}/d\lambda. \tag{8.18}$$

In a local Lorentz frame, the components of $\boldsymbol{\nabla T}$ are directional derivatives of the components of $\boldsymbol{T}$: $T^{\beta}{}_{\alpha,\gamma}$. Not so in a general basis. If $\{\boldsymbol{e}_{\beta}(\mathscr{P})\}$ is a basis that varies arbitrarily but smoothly from point to point, and $\{\boldsymbol{\omega}^{\alpha}(\mathscr{P})\}$ is its dual basis, then $\boldsymbol{\nabla T} = \boldsymbol{\nabla}(T^{\beta}{}_{\alpha}\boldsymbol{e}_{\beta} \otimes \boldsymbol{\omega}^{\alpha})$ will contain contributions from $\boldsymbol{\nabla e}_{\beta}$ and $\boldsymbol{\nabla\omega}^{\alpha}$, as well as from $\boldsymbol{\nabla} T^{\beta}{}_{\alpha} \equiv \boldsymbol{d}T^{\beta}{}_{\alpha} = T^{\beta}{}_{\alpha,\gamma}\boldsymbol{\omega}^{\gamma}$.

Connection coefficients defined

To quantify the contributions from $\boldsymbol{\nabla e}_{\beta}$ and $\boldsymbol{\nabla\omega}^{\alpha}$, i.e., to quantify the twisting, turning, expansion, and contraction of the basis vectors and 1-forms, one defines *"connection coefficients"*:

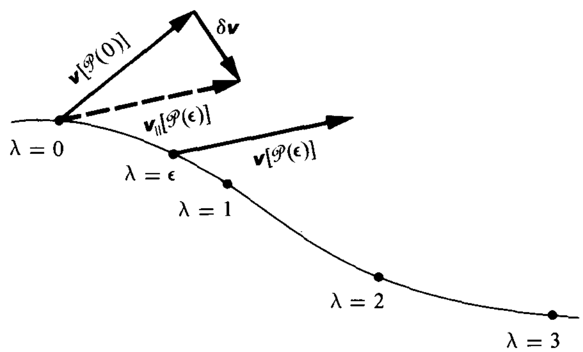


Figure 8.2.
Definition of the covariant derivative "$\nabla_{u} v$" of a vector field v along a curve $\mathcal{P}(\lambda)$, with tangent vector $u \equiv d\mathcal{P}/d\lambda$: (1) choose a point $\mathcal{P}(0)$ on the curve, at which to evaluate $\nabla_u v$. (2) Choose a nearby point $\mathcal{P}(\varepsilon)$ on the curve. (3) Parallel-transport $v[\mathcal{P}(\varepsilon)]$ along the curve back to $\mathcal{P}(0)$, getting the vector $v_{\parallel}[\mathcal{P}(\varepsilon)]$. (4) Take the difference $\delta v \equiv v_{\parallel}[\mathcal{P}(\varepsilon)] - v[\mathcal{P}(0)]$. (5) Then $\nabla_u v$ is defined by

$$\nabla_u v \equiv \lim_{\varepsilon \to 0} \frac{\delta v}{\varepsilon} = \lim_{\varepsilon \to 0} \left\{ \frac{v_{\parallel}[\mathcal{P}(\varepsilon)] - v[\mathcal{P}(0)]}{\varepsilon} \right\}.$$

$$\Gamma^{\alpha}{}_{\beta\gamma} \equiv \langle \boldsymbol{\omega}^{\alpha}, \underbrace{\nabla_{\gamma}}_{[\equiv \nabla_{e_\gamma}]} \boldsymbol{e}_{\beta} \rangle \qquad \begin{bmatrix} \text{Note reversal of } \beta \text{ and } \gamma \text{ to make the} \\ \text{differentiating index come last on } \Gamma \end{bmatrix} \tag{8.19a}$$

$$= \begin{pmatrix} \alpha \text{ component of change in } \boldsymbol{e}_{\beta}, \text{ relative} \\ \text{to parallel transport, along } \boldsymbol{e}_{\gamma} \end{pmatrix},$$

and one proves (exercise 8.12) that

$$\langle \nabla_{\gamma} \boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta} \rangle = -\Gamma^{\alpha}{}_{\beta\gamma}. \tag{8.19b}$$

In terms of these coefficients and

$$T^{\beta}{}_{\alpha,\gamma} \equiv \nabla_{\gamma} T^{\beta}{}_{\alpha} \equiv \partial_{e_\gamma} T^{\beta}{}_{\alpha} \equiv \partial_{\gamma} T^{\beta}{}_{\alpha}, \tag{8.20}$$

the components of the gradient, denoted $T^{\beta}{}_{\alpha;\gamma}$, are

Components of gradient in arbitrary frame

$$T^{\beta}{}_{\alpha;\gamma} = T^{\beta}{}_{\alpha,\gamma} + \Gamma^{\beta}{}_{\mu\gamma} T^{\mu}{}_{\alpha} - \Gamma^{\mu}{}_{\alpha\gamma} T^{\beta}{}_{\mu} \tag{8.21}$$

(see exercise 8.13). If the basis at the event where $\nabla \boldsymbol{T}$ is calculated were a local Lorentz frame, the components of $\nabla \boldsymbol{T}$ would just be $T^{\beta}{}_{\alpha,\gamma}$. Because it is not, one must correct this "Lorentz-frame" value for the twisting, turning, expansion, and contraction of the basis vectors and 1-forms. The "ΓT" terms in equation (8.21) are the necessary corrections—one for each index of $\boldsymbol{T}$. The pattern of these correction terms is easy to remember: (1) "+" sign if index being corrected is up, "−" sign if it is down; (2) differentiation index (γ in above case) always at end of Γ; (3) index being corrected (β in first term, α in second) shifts from T onto Γ and gets replaced on T by a dummy summation index (μ).

Knowing the components (8.21) of the gradient, one can calculate the components of the covariant derivative $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{T}$ by a simple contraction into u^γ [see equation (8.17)]:

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{T} = (T^\beta{}_{\alpha;\gamma}u^\gamma)\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha. \tag{8.22}$$

Components of covariant derivative

When $\boldsymbol{u}$ is the tangent vector to a curve $\mathscr{P}(\lambda)$, $\boldsymbol{u} = d\mathscr{P}/d\lambda$, one uses the notation $DT^\beta{}_\alpha/d\lambda$ for the components of $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{T}$:

$$\begin{aligned}\frac{DT^\beta{}_\alpha}{d\lambda} &\equiv T^\beta{}_{\alpha;\gamma}u^\gamma \underset{\text{[if basis is a coordinate basis so } u^\gamma = dx^\gamma/d\lambda]}{=} T^\beta{}_{\alpha;\gamma}\frac{dx^\gamma}{d\lambda} \\ &= (T^\beta{}_{\alpha,\gamma} + \text{“}\Gamma T\text{” corrections})\, dx^\gamma/d\lambda \\ &= \frac{dT^\beta{}_\alpha}{d\lambda} + (\Gamma^\beta{}_{\mu\gamma}T^\mu{}_\alpha - \Gamma^\mu{}_{\alpha\gamma}T^\beta{}_\mu)\frac{dx^\gamma}{d\lambda}.\end{aligned} \tag{8.23}$$

The “;” in $T^\beta{}_{\alpha;\gamma}$ reminds one to correct $T^\beta{}_{\alpha,\gamma}$ with “ΓT” terms; similarly, the “D” in $DT^\beta{}_\alpha/d\lambda$ reminds one to correct $dT^\beta{}_\alpha/d\lambda$ with “ΓT” terms.

Calculation of connection coefficients from metric and commutators

This is all well and good, but how does one find out the connection coefficients $\Gamma^\alpha{}_{\beta\gamma}$ for a given basis? The answer is derived in exercise 8.15. It says: (1) take the metric coefficients in the given basis; (2) calculate their directional derivatives along the basis directions

$$g_{\beta\gamma,\mu} \equiv \underset{[\equiv \partial_{\boldsymbol{e}_\mu}]}{\partial_\mu g_{\beta\gamma}} \underset{\text{[if a coordinate basis, } \boldsymbol{e}_\mu = \partial\mathscr{P}/\partial x^\mu\text{, is being used]}}{=} \partial g_{\beta\gamma}/\partial x^\mu; \tag{8.24a}$$

(3) calculate the commutation coefficients of the basis [equations (8.14) in general; $c_{\mu\beta\gamma} = 0$ in special case of coordinate basis]; (4) calculate the “covariant connection coefficients”

$$\Gamma_{\mu\beta\gamma} = \frac{1}{2}(g_{\mu\beta,\gamma} + g_{\mu\gamma,\beta} - g_{\beta\gamma,\mu} + \underbrace{c_{\mu\beta\gamma} + c_{\mu\gamma\beta} - c_{\beta\gamma\mu}}_{\text{these terms are 0 for coordinate basis}}); \tag{8.24b}$$

(5) raise an index to get the connection coefficients:

$$\Gamma^\alpha{}_{\beta\gamma} = g^{\alpha\mu}\Gamma_{\mu\beta\gamma}. \tag{8.24c}$$

[*Note on terminology:* a coordinate basis always has $c_{\alpha\beta\gamma} = 0$, and is sometimes called *holonomic;* a noncoordinate basis always has some of its $c_{\alpha\beta\gamma}$ nonzero, and is sometimes called *anholonomic*. In the holonomic case, the connection coefficients are sometimes called *Christoffel symbols*.]

The component notation, with its semicolons, commas, D's, connection coefficients, etc., looks rather formidable at first. But it bears great computational power, one discovers as one proceeds deep into gravitation theory; and its rules of manipulation are simple enough to be learned easily. By contrast, the abstract notation ($\boldsymbol{\nabla} \boldsymbol{T}$, $\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{T}$, etc.) is poorly suited to complex calculations; but it possesses great conceptual power.

This contrast shows clearly in the way the two notations handle the concept of *geodesic*. A geodesic of spacetime is a curve that is straight and uniformly parametrized, as measured in each local Lorentz frame along its way. If the geodesic is timelike, then it is a possible world line for a freely falling particle, and its uniformly ticking parameter λ (called "*affine parameter*") is a multiple of the particle's proper time, $\lambda = a\tau + b$. (Principle of equivalence: test particles move on straight lines in local Lorentz frames, and each particle's clock ticks at a uniform rate as measured by any Lorentz observer.) This definition of geodesic is readily translated into abstract, coordinate-free language: a geodesic is a curve $\mathscr{P}(\lambda)$ that parallel-transports its tangent vector $\boldsymbol{u} = d\mathscr{P}/d\lambda$ along itself—

Geodesic and affine parameter defined

$$\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{u} = 0. \tag{8.25}$$

(See Figure 10.1.) What could be simpler conceptually? But to compute the geodesic, given an initial event $\mathscr{P}_0$ and initial tangent vector $\boldsymbol{u}(\mathscr{P}_0)$ there, one must use the component formalism. Introduce a coordinate system $x^\alpha(\mathscr{P})$, in which $u^\alpha = dx^\alpha/d\lambda$, and write the component version of equation (8.25) as

$$0 = \frac{D(dx^\alpha/d\lambda)}{d\lambda} = \frac{d(dx^\alpha/d\lambda)}{d\lambda} + \left(\Gamma^\alpha{}_{\mu\gamma}\frac{dx^\mu}{d\lambda}\right)\frac{dx^\gamma}{d\lambda}$$

[see equation (8.23), with one less index on T]; i.e.,

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha{}_{\mu\gamma}\frac{dx^\mu}{d\lambda}\frac{dx^\gamma}{d\lambda} = 0. \tag{8.26}$$

Geodesic equation

This *geodesic equation* can be solved (in principle) for the coordinates of the geodesic, $x^\alpha(\lambda)$, when initial data [x^α and $dx^\alpha/d\lambda$ at $\lambda = \lambda_0$] have been specified.

The geodesics of the Earth's surface (great circles) are a foil against which one can visualize connection coefficients; see Figure 8.3.

The material of this section is presented more deeply and from a different viewpoint in Chapters 10 and 13. The Track-2 reader who plans to study those chapters is advised to ignore the following exercises. The Track-1 reader who intends to skip Chapters 9–15 will gain necessary experience with the component formalism by working exercises 8.4–8.7. Less important to him, but valuable nonetheless, are exercises 8.8–8.15, which develop the formalism of covariant derivatives and connection coefficients in a systematic manner. The most important results of these exercises will be summarized in Box 8.6 (pages 223 and 224).

EXERCISES

Exercise 8.4. PRACTICE IN WRITING COMPONENTS OF GRADIENT

Rewrite the following quantities in terms of ordinary derivatives ($f_{,\gamma} \equiv \partial_{\boldsymbol{e}_\gamma} f \equiv \boldsymbol{\nabla}_\gamma f$) and "$\Gamma T$" correction terms: (a) $T_{;\gamma}$ where T is a function. (b) $T^\alpha{}_{;\gamma}$ where $\boldsymbol{T}$ is a vector. (c) $T_{\alpha;\gamma}$ where $\boldsymbol{T}$ is a 1-form. (d) $T^\alpha{}_{\beta\delta}{}^\varepsilon{}_{;\gamma}$. [*Answer:*

(a) $T_{;\gamma} = T_{,\gamma}$. (b) $T^\alpha{}_{;\gamma} = T^\alpha{}_{,\gamma} + \Gamma^\alpha{}_{\mu\gamma}T^\mu$. (c) $T_{\alpha;\gamma} = T_{\alpha,\gamma} - \Gamma^\mu{}_{\alpha\gamma}T_\mu$.

(d) $T^\alpha{}_{\beta\delta}{}^\varepsilon{}_{;\gamma} = T^\alpha{}_{\beta\delta}{}^\varepsilon{}_{,\gamma} + \Gamma^\alpha{}_{\mu\gamma}T^\mu{}_{\beta\delta}{}^\varepsilon - \Gamma^\mu{}_{\beta\gamma}T^\alpha{}_{\mu\delta}{}^\varepsilon - \Gamma^\mu{}_{\delta\gamma}T^\alpha{}_{\beta\mu}{}^\varepsilon + \Gamma^\varepsilon{}_{\mu\gamma}T^\alpha{}_{\beta\delta}{}^\mu$.]

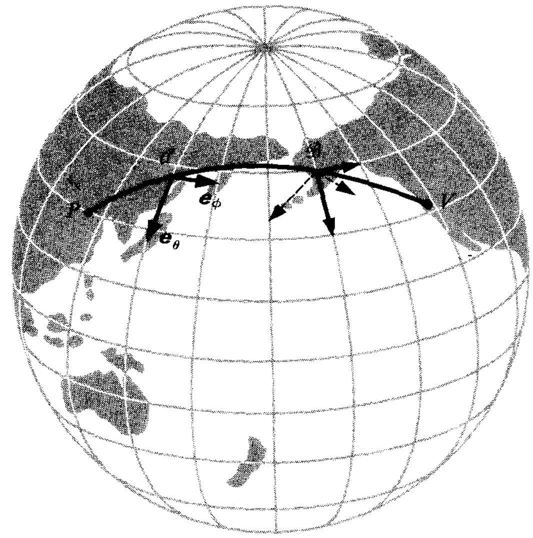


Figure 8.3.
The why of connection coefficients, schematically portrayed. The aviator pursuing his great circle route from Peking to Vancouver finds himself early going north, but later going south, although he is navigating the straightest route that is at all open to him (geodesic). The apparent change in direction indicates a turning, not in his route, but in the system of coordinates with respect to which his route is described. The vector $\boldsymbol{v}$ of his velocity (a vector defined not on spacetime but rather on the Earth's two-dimensional surface), carried forward by parallel transport from an earlier moment to a later moment, finds itself in agreement with the velocity that he is then pursuing; or, in the abstract language of coordinate-free differential geometry, the covariant derivative $\nabla_{\boldsymbol{v}}\boldsymbol{v}$ vanishes along the route ("equation of a geodesic"). Though $\boldsymbol{v}$ is in this sense constant, the individual pieces of which the navigator considers this vector to be built, $\boldsymbol{v} = v^\theta \boldsymbol{e}_\theta + v^\phi \boldsymbol{e}_\phi$, are not constant.

In the language of components, the quantities v^θ and v^ϕ are changing along the route at a rate that annuls the covariant derivative of $\boldsymbol{v}$; thus

$$\nabla_{\boldsymbol{v}}\boldsymbol{v} = \boldsymbol{a} = a^\phi \boldsymbol{e}_\phi + a^\theta \boldsymbol{e}_\theta = 0,$$

or

$$0 = a^\theta = \frac{dv^\theta}{dt} + \Gamma^\theta{}_{mn} v^m v^n,$$

$$0 = a^\phi = \frac{dv^\phi}{dt} + \Gamma^\phi{}_{mn} v^m v^n.$$

In this sense the connection coefficients $\Gamma^j{}_{mn}$ serve as "turning coefficients" to tell how fast to "turn" the components of a vector in order to keep that vector constant (against the turning influence of the base vectors).

Alternatively, the navigator can use an "automatic pilot system" which parallel-transports its own base vectors along the plane's route:

$$\nabla_{\boldsymbol{v}}\boldsymbol{e}_{\theta'} = \nabla_{\boldsymbol{v}}\boldsymbol{e}_{\phi'} = 0;$$

solid vectors at $\mathcal{A}$ become dotted vectors at $\mathcal{B}$. Then the components of $\boldsymbol{v}$ must be kept fixed to achieve a great-circle route,

$$\frac{dv^{\theta'}}{dt} = \frac{dv^{\phi'}}{dt} = 0;$$

and the turning coefficients are used to describe the turning of the lines of latitude and longitude relative to this parallel-transported basis:

$$\nabla_{\boldsymbol{v}}\boldsymbol{e}_\theta = \boldsymbol{e}_m \Gamma^m{}_{\theta n} v^n,$$
$$\nabla_{\boldsymbol{v}}\boldsymbol{e}_\phi = \boldsymbol{e}_m \Gamma^m{}_{\phi n} v^n.$$

The same turning coefficients enter into both viewpoints. The only difference is in how these coefficients are used.

Exercise 8.5. A SHEET OF PAPER IN POLAR COORDINATES

The two-dimensional metric for a flat sheet of paper in polar coordinates (r, θ) is $ds^2 = dr^2 + r^2\, d\phi^2$—or, in modern notation, $\boldsymbol{g} = \boldsymbol{dr} \otimes \boldsymbol{dr} + r^2\, \boldsymbol{d\phi} \otimes \boldsymbol{d\phi}$.

(a) Calculate the connection coefficients using equations (8.24). [*Answer:* $\Gamma^r{}_{\phi\phi} = -r$; $\Gamma^\phi{}_{r\phi} = \Gamma^\phi{}_{\phi r} = 1/r$; all others vanish.]

(b) Write down the geodesic equation in (r, ϕ) coordinates. [*Answer:* $d^2r/d\lambda^2 - r(d\phi/d\lambda)^2 = 0$; $d^2\phi/d\lambda^2 + (2/r)(dr/d\lambda)(d\phi/d\lambda) = 0$.]

(c) Solve this geodesic equation for $r(\lambda)$ and $\phi(\lambda)$, and show that the solution is a uniformly parametrized straight line ($x \equiv r\cos\phi = a\lambda + b$ for some a and b; $y \equiv r \sin\phi = j\lambda + k$ for some j and k).

(d) Verify that the noncoordinate basis $\boldsymbol{e}_{\hat r} \equiv \boldsymbol{e}_r = \partial\mathcal{P}/\partial r$, $\boldsymbol{e}_{\hat\phi} \equiv r^{-1}\boldsymbol{e}_\phi \quad r^{-1}\,\partial\mathcal{P}/\partial\phi$, $\boldsymbol{\omega}^{\hat r} = \boldsymbol{dr}$, $\boldsymbol{\omega}^{\hat\phi} = r\,\boldsymbol{d\phi}$ is orthonormal, and that $\langle \boldsymbol{\omega}^{\hat\alpha}, \boldsymbol{e}_{\hat\beta}\rangle = \delta^{\hat\alpha}{}_{\hat\beta}$. Then calculate the connection coefficients of this basis from a knowledge [part (a)] of the connection of the coordinate basis. [*Answer:*

$$
\begin{aligned}
\Gamma^{\hat\phi}{}_{\hat\phi\hat r} &= \langle \boldsymbol{\omega}^{\hat\phi}, \boldsymbol{\nabla}_{\hat r}\boldsymbol{e}_{\hat\phi}\rangle = \langle r\,\boldsymbol{d\phi}, \boldsymbol{\nabla}_r(r^{-1}\boldsymbol{e}_\phi)\rangle \\
&= r\langle \boldsymbol{d\phi}, (\boldsymbol{\nabla}_r r^{-1})\boldsymbol{e}_\phi + r^{-1}(\boldsymbol{\nabla}_r\boldsymbol{e}_\phi)\rangle = r\langle \boldsymbol{d\phi}, -r^{-2}\boldsymbol{e}_\phi\rangle + \langle \boldsymbol{d\phi}, \boldsymbol{\nabla}_r\boldsymbol{e}_\phi\rangle \\
&= -r^{-1} + \Gamma^\phi{}_{\phi r} = -r^{-1} + r^{-1} = 0;
\end{aligned}
$$

similarly, $\Gamma^{\hat\phi}{}_{\hat r\hat\phi} = +1/r$, $\Gamma^{\hat r}{}_{\hat\phi\hat\phi} = -1/r$; all others vanish.]

(e) Consider the Keplerian orbit of Figure 8.1 and §8.3 as a nongeodesic curve in the sun's two-dimensional, Euclidean, equatorial plane. In place of the old notation $d\boldsymbol{v}/dt$, $d\boldsymbol{e}_{\hat r}/dt$, etc., use the new notation $\boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{v}$, $\boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{e}_{\hat r}$, etc. Then $\boldsymbol{v} = d\mathcal{P}/dt$ is the tangent to the orbit, and $\boldsymbol{a} = \boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{v}$ is the acceleration. Derive equations (8.4) for $a^{\hat r}$ and $a^{\hat\phi}$ using component manipulations and connection coefficients in the orthonormal basis.

Exercise 8.6. SPHERICAL COORDINATES IN FLAT SPACETIME

The spherical noncoordinate basis $\{\boldsymbol{e}_{\hat\alpha}\}$ of Exercise 8.1 was orthonormal, $g_{\hat\alpha\hat\beta} = \eta_{\hat\alpha\hat\beta}$, but had nonvanishing commutation coefficients [part (c) of Exercise 8.2].

(a) Calculate the connection coefficients for this basis, using equations (8.24). [*Answer:*

$$
\begin{aligned}
&\Gamma^{\hat\theta}{}_{\hat r\hat\theta} = \Gamma^{\hat\phi}{}_{\hat r\hat\phi} = -\Gamma^{\hat r}{}_{\hat\theta\hat\theta} = -\Gamma^{\hat r}{}_{\hat\phi\hat\phi} = 1/r;\\
&\Gamma^{\hat\phi}{}_{\hat\theta\hat\phi} = -\Gamma^{\hat\theta}{}_{\hat\phi\hat\phi} = \cot\theta/r;
\end{aligned}
$$

all others vanish.]

(b) Write down expressions for $\boldsymbol{\nabla}_{\hat\alpha}\boldsymbol{e}_{\hat\beta}$ in terms of $\boldsymbol{e}_{\hat\gamma}$, and verify the correctness of these expressions by drawing sketches of the basis vectors on a sphere of constant t and r. [*Answer:*

$$
\boldsymbol{\nabla}_{\hat\theta}\boldsymbol{e}_{\hat r} = r^{-1}\boldsymbol{e}_{\hat\theta}, \qquad \boldsymbol{\nabla}_{\hat\theta}\boldsymbol{e}_{\hat\theta} = -r^{-1}\boldsymbol{e}_{\hat r}, \qquad \boldsymbol{\nabla}_{\hat\phi}\boldsymbol{e}_{\hat r} = r^{-1}\boldsymbol{e}_{\hat\phi},
$$

$$
\boldsymbol{\nabla}_{\hat\phi}\boldsymbol{e}_{\hat\theta} = (\cot\theta/r)\boldsymbol{e}_{\hat\phi}, \qquad \boldsymbol{\nabla}_{\hat\phi}\boldsymbol{e}_{\hat\phi} = -r^{-1}\boldsymbol{e}_{\hat r} - (\cot\theta/r)\boldsymbol{e}_{\hat\theta}.
$$

All others vanish.]

(c) Calculate the divergence of a vector, $\boldsymbol{\nabla}\cdot\boldsymbol{A} = A^{\hat\alpha}{}_{;\hat\alpha}$, in this basis. [*Answer:*

$$
\begin{aligned}
\boldsymbol{\nabla}\cdot\boldsymbol{A} &= A^{\hat t}{}_{,\hat t} + r^{-2}(r^2A^{\hat r})_{,\hat r} + (\sin\theta)^{-1}(\sin\theta A^{\hat\theta})_{,\hat\theta} + A^{\hat\phi}{}_{,\hat\phi} \\
&= \frac{\partial A^{\hat t}}{\partial t} + \frac{1}{r^2}\frac{\partial(r^2A^{\hat r})}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta A^{\hat\theta})}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial A^{\hat\phi}}{\partial\phi}.
\end{aligned}
$$

This answer should be familiar from flat-space vector analysis.]

Exercise 8.7. SYMMETRIES OF CONNECTION COEFFICIENTS

From equation (8.24b), the symmetry of the metric, and the antisymmetry ($c_{\beta\gamma\mu} = -c_{\gamma\beta\mu}$)

of the commutation coefficients, show that: $\Gamma_{\alpha[\beta\gamma]} = 0$ (last two indices are symmetric) in a coordinate basis; $\Gamma_{(\hat{\alpha}\hat{\beta})\hat{\gamma}} = 0$ (first two indices are antisymmetric) in a globally orthonormal basis, $g_{\hat{\alpha}\hat{\beta}} = \eta_{\alpha\beta}$.

SYSTEMATIC DERIVATION OF RESULTS IN §8.5

Exercise 8.8. NEW DEFINITION OF $\nabla \boldsymbol{T}$ COMPARED WITH OLD DEFINITION

The new definition of $\nabla \boldsymbol{T}$ is given by equations (8.16) and (8.17). Use the fact that parallel transport keeps local-Lorentz components fixed to derive, from (8.16), the Lorentz-frame equation $\nabla_{\boldsymbol{u}} \boldsymbol{T} = T^{\beta}{}_{\alpha,\gamma} u^{\gamma} \boldsymbol{e}_{\beta} \otimes \boldsymbol{\omega}^{\alpha}$. From this and equation (8.17), infer that the Lorentz-frame components of $\nabla \boldsymbol{T}$ are $T^{\beta}{}_{\alpha,\gamma}$—which accords with the old definition of $\nabla \boldsymbol{T}$.

Exercise 8.9. CHAIN RULE FOR $\nabla_{\boldsymbol{u}} \boldsymbol{T}$

(a) Use calculations in a local Lorentz frame to show that "$\nabla_{\boldsymbol{u}}$" obeys the standard chain rule for derivatives:

$$\nabla_{\boldsymbol{u}}(f\boldsymbol{A} \otimes \boldsymbol{B}) = (\nabla_{\boldsymbol{u}} f)\boldsymbol{A} \otimes \boldsymbol{B} + f(\nabla_{\boldsymbol{u}} \boldsymbol{A}) \otimes \boldsymbol{B} + f\boldsymbol{A} \otimes (\nabla_{\boldsymbol{u}} \boldsymbol{B}). \tag{8.27}$$

Here $\boldsymbol{A}$ and $\boldsymbol{B}$ are arbitrary vectors, 1-forms, or tensors; and f is an arbitrary function. [*Hint:* assume for concreteness that $\boldsymbol{A}$ is a $\binom{1}{1}$ tensor and $\boldsymbol{B}$ is a vector. Then this equation reads, in Lorentz-frame component notation,

$$(fA^{\alpha}{}_{\beta}B^{\gamma})_{,\delta}u^{\delta} = (f_{,\delta}u^{\delta})A^{\alpha}{}_{\beta}B^{\gamma} + f(A^{\alpha}{}_{\beta,\delta}u^{\delta})B^{\gamma} + fA^{\alpha}{}_{\beta}(B^{\gamma}{}_{,\delta}u^{\delta}).] \tag{8.27'}$$

(b) Rewrite equation (8.27) in component notation in an arbitrary basis. [*Answer:* same as (8.27′), except "," is replaced everywhere by ";". But *note* that $f_{;\delta}u^{\delta} = f_{,\delta}u^{\delta}$, because the function f "has no components to correct".]

Exercise 8.10. COVARIANT DERIVATIVE COMMUTES WITH CONTRACTION

(a) Let $\boldsymbol{S}$ be a $\binom{1}{2}$ tensor. Using components in a local Lorentz frame show that

$$\nabla_{\boldsymbol{u}}\,(\text{contraction on slots 1 and 2 of } \boldsymbol{S}) = (\text{contraction on slots 1 and 2 of } \nabla_{\boldsymbol{u}}\boldsymbol{S}). \tag{8.28}$$

[Hint: in a local Lorentz frame this equation makes the trivial statement

$$\left(\sum_{\alpha} S^{\alpha}{}_{\alpha\beta}\right)_{,\gamma} u^{\gamma} = \sum_{\alpha} (S^{\alpha}{}_{\alpha\beta,\gamma}u^{\gamma}).]$$

Exercise 8.11. ALGEBRAIC PROPERTIES OF ∇

Use calculations in a local Lorentz frame to show that

$$\nabla_{a\boldsymbol{u}+b\boldsymbol{v}}\boldsymbol{S} = a\,\nabla_{\boldsymbol{u}}\boldsymbol{S} + b\,\nabla_{\boldsymbol{v}}\boldsymbol{S} \tag{8.29}$$

for all tangent vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and numbers a, b; also that

$$\nabla_{\boldsymbol{u}}(\boldsymbol{S} + \boldsymbol{M}) = \nabla_{\boldsymbol{u}}\boldsymbol{S} + \nabla_{\boldsymbol{u}}\boldsymbol{M} \tag{8.30}$$

for any two tensor fields $\boldsymbol{S}$ and $\boldsymbol{M}$ of the same rank; also that

$$\nabla_{\boldsymbol{u}}\boldsymbol{w} - \nabla_{\boldsymbol{w}}\boldsymbol{u} = \underbrace{[\boldsymbol{u}, \boldsymbol{w}]}_{\left[\substack{\text{commutator of } \boldsymbol{u} \text{ and } \boldsymbol{w};\\ \text{discussed in exercise 8.2}}\right]}, \tag{8.31}$$

for any two vector fields $\boldsymbol{u}$ and $\boldsymbol{w}$.

Exercise 8.12. CONNECTION COEFFICIENTS FOR 1-FORM BASIS

Show that the same connection coefficients $\Gamma^{\alpha}{}_{\beta\gamma}$ that describe the changes in $\{\boldsymbol{e}_\beta\}$ from point to point [definition (8.19a)] also describe the changes in $\{\boldsymbol{\omega}^\alpha\}$, except for a change in sign [equation (8.19b)]. {*Answer:* (1) $\langle\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta\rangle = \delta^\alpha{}_\beta$ is a constant function (0 or 1, depending on whether $\alpha = \beta$). (2) Thus, $\nabla_\gamma\langle\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta\rangle = \partial_{\boldsymbol{e}_\gamma}\langle\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta\rangle = 0$. (3) But $\langle\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta\rangle$ is the contraction of $\boldsymbol{\omega}^\alpha \otimes \boldsymbol{e}_\beta$, so equation (8.28) implies $0 = \nabla_\gamma$(contraction of $\boldsymbol{\omega}^\alpha \otimes \boldsymbol{e}_\beta$) $=$ contraction of $[\nabla_\gamma(\boldsymbol{\omega}^\alpha \otimes \boldsymbol{e}_\beta)]$. (4) Apply the chain rule (8.27) to conclude $0 =$ contraction of $[(\nabla_\gamma\boldsymbol{\omega}^\alpha) \otimes \boldsymbol{e}_\beta + \boldsymbol{\omega}^\alpha \otimes (\nabla_\gamma\boldsymbol{e}_\beta)] = \langle\nabla_\gamma\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta\rangle + \langle\boldsymbol{\omega}^\alpha, \nabla_\gamma\boldsymbol{e}_\beta\rangle$. (5) Finally, use definition (8.19a) to arrive at the desired result, (8.19b).}

Exercise 8.13. "ΓT" CORRECTION TERMS FOR $T^{\beta}{}_{\alpha;\gamma}$

Derive equation (8.21) for $T^\beta{}_{\alpha;\gamma}$ in an arbitrary basis by first calculating the components of $\nabla_{\boldsymbol{u}}\boldsymbol{T}$ for arbitrary $\boldsymbol{u}$, and by then using equation (8.17) to infer the components of $\nabla\boldsymbol{T}$. [*Answer:* (1) Use the chain rule (8.27) to get

$$\begin{aligned}\nabla_{\boldsymbol{u}}\boldsymbol{T} &= \nabla_{\boldsymbol{u}}(T^\beta{}_\alpha\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha)\\ &= (\nabla_{\boldsymbol{u}}T^\beta{}_\alpha)\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha + T^\beta{}_\alpha(\nabla_{\boldsymbol{u}}\boldsymbol{e}_\beta) \otimes \boldsymbol{\omega}^\alpha + T^\beta{}_\alpha\boldsymbol{e}_\beta \otimes (\nabla_{\boldsymbol{u}}\boldsymbol{\omega}^\alpha).\end{aligned}$$

(2) Write $\boldsymbol{u}$ in terms of its components, $\boldsymbol{u} = u^\gamma\boldsymbol{e}_\gamma$; use linearity of $\nabla_{\boldsymbol{u}}$ in $\boldsymbol{u}$ from equation (8.29), to get $\nabla_{\boldsymbol{u}} = u^\gamma\nabla_\gamma$; and use this in $\nabla_{\boldsymbol{u}}\boldsymbol{T}$:

$$\nabla_{\boldsymbol{u}}\boldsymbol{T} = u^\gamma\{T^\beta{}_{\alpha,\gamma}\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha + T^\beta{}_\alpha(\nabla_\gamma\boldsymbol{e}_\beta) \otimes \boldsymbol{\omega}^\alpha + T^\beta{}_\alpha\boldsymbol{e}_\beta \otimes (\nabla_\gamma\boldsymbol{\omega}^\alpha)\}.$$

(3) Use equations (8.19a,b), rewritten as

$$\nabla_\gamma\boldsymbol{e}_\beta = \Gamma^\mu{}_{\beta\gamma}\boldsymbol{e}_\mu, \qquad \nabla_\gamma\boldsymbol{\omega}^\alpha = -\Gamma^\alpha{}_{\mu\gamma}\boldsymbol{\omega}^\mu, \tag{8.32}$$

to put $\nabla_{\boldsymbol{u}}\boldsymbol{T}$ in the form

$$\nabla_{\boldsymbol{u}}\boldsymbol{T} = u^\gamma\{T^\beta{}_{\alpha,\gamma}\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha + \Gamma^\mu{}_{\beta\gamma}T^\beta{}_\alpha\boldsymbol{e}_\mu \otimes \boldsymbol{\omega}^\alpha - \Gamma^\alpha{}_{\mu\gamma}T^\beta{}_\alpha\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\mu\}.$$

(4) Rename dummy indices so that the basis tensor $\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha$ can be factored out:

$$\nabla_{\boldsymbol{u}}\boldsymbol{T} = u^\gamma\{T^\beta{}_{\alpha,\gamma} + \Gamma^\beta{}_{\mu\gamma}T^\mu{}_\alpha - \Gamma^\mu{}_{\alpha\gamma}T^\beta{}_\mu\}\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha.$$

(5) By comparison with

$$\nabla_{\boldsymbol{u}}\boldsymbol{T} = \nabla\boldsymbol{T}(\ldots, \ldots, \boldsymbol{u}) = (T^\beta{}_{\alpha;\gamma}u^\gamma)\boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\alpha,$$

read off the value of $T^\beta{}_{\alpha;\gamma}$.]

Exercise 8.14. METRIC IS COVARIANTLY CONSTANT

Show on physical grounds (using properties of local Lorentz frames) that

$$\nabla\boldsymbol{g} = 0 \tag{8.33}$$

or, equivalently, that $\nabla_{\boldsymbol{u}} \boldsymbol{g} = 0$ for any vector $\boldsymbol{u}$. Then deduce as a mathematical consequence the obviously desirable product rule

$$\nabla_{\boldsymbol{u}}(\boldsymbol{A} \cdot \boldsymbol{B}) = (\nabla_{\boldsymbol{u}} \boldsymbol{A}) \cdot \boldsymbol{B} + \boldsymbol{A} \cdot (\nabla_{\boldsymbol{u}} \boldsymbol{B}).$$

[Answer: (1) As discussed following equation (8.18), the components of $\nabla \boldsymbol{g}$ in a local Lorentz frame are $g_{\mu\nu,\alpha}$. Just use $\boldsymbol{g}$ for $\boldsymbol{T}$ in that discussion. But these components all vanish by equation (8.15b). Therefore equation (8.33) holds in this frame, and—as a tensor equation—in all frames. (2) The product rule is also a tensor equation, true immediately via components in a local Lorentz frame. (3) Prove the product rule also the hard way, to see where equation (8.33) enters. Use the chain rule of exercise 8.9 to write

$$\begin{aligned}\nabla_{\boldsymbol{u}}(\boldsymbol{g} \otimes \boldsymbol{A} \otimes \boldsymbol{B}) = (\nabla_{\boldsymbol{u}} \boldsymbol{g}) \otimes \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{g} \otimes (\nabla_{\boldsymbol{u}} \boldsymbol{A}) \otimes \boldsymbol{B} \\ + \boldsymbol{g} \otimes \boldsymbol{A} \otimes (\nabla_{\boldsymbol{u}} \boldsymbol{B}).\end{aligned}$$

Use equation (8.33) to drop one term, then contract, forming

$$\boldsymbol{A} \cdot \boldsymbol{B} = \text{contraction}\,(\boldsymbol{g} \otimes \boldsymbol{A} \otimes \boldsymbol{B})$$

and the other inner products. Exercise 8.10 is used to justify commuting the contraction with $\nabla_{\boldsymbol{u}}$ on the lefthand side.]

Exercise 8.15. CONNECTION COEFFICIENTS IN TERMS OF METRIC

Use the fact that the metric is covariantly constant [equation (8.33)] to derive equation (8.24b) for the connection coefficients. Treat equation (8.24c) as a definition of $\Gamma_{\mu\beta\gamma}$ in terms of $\Gamma^{\alpha}{}_{\beta\gamma}$. [*Answer:* (1) Calculate the components of $\nabla \boldsymbol{g}$ in an arbitrary frame:

$$\begin{aligned}g_{\alpha\beta;\gamma} = 0 &= g_{\alpha\beta,\gamma} - \Gamma^{\mu}{}_{\alpha\gamma} g_{\mu\beta} - \Gamma^{\mu}{}_{\beta\gamma} g_{\mu\alpha} \\ &\equiv g_{\alpha\beta,\gamma} - \Gamma_{\beta\alpha\gamma} - \Gamma_{\alpha\beta\gamma};\end{aligned}$$

thereby conclude that $g_{\alpha\beta,\gamma} = 2\Gamma_{(\alpha\beta)\gamma}$. (Round brackets denote symmetric part.) (2) Construct the metric terms in the claimed answer for $\Gamma_{\mu\beta\gamma}$:

$$\begin{aligned}\frac{1}{2}(g_{\mu\beta,\gamma} + g_{\mu\gamma,\beta} - g_{\beta\gamma,\mu}) &= \Gamma_{(\mu\beta)\gamma} + \Gamma_{(\mu\gamma)\beta} - \Gamma_{(\beta\gamma)\mu} \\ &= \frac{1}{2}[\Gamma_{\mu\beta\gamma} + \Gamma_{\beta\mu\gamma} + \Gamma_{\mu\gamma\beta} + \Gamma_{\gamma\mu\beta} - \Gamma_{\beta\gamma\mu} - \Gamma_{\gamma\beta\mu}] \\ &= \Gamma_{\mu\beta\gamma} + (-\Gamma_{\mu[\beta\gamma]} + \Gamma_{\beta[\mu\gamma]} + \Gamma_{\gamma[\mu\beta]}).\end{aligned}$$

(3) Infer from equation (8.31), with $\boldsymbol{u}$ and $\boldsymbol{w}$ chosen as two basis vectors ($\boldsymbol{u} = \boldsymbol{e}_\mu$, $\boldsymbol{w} = \boldsymbol{e}_\nu$) that

$$c_{\mu\nu}{}^{\rho}\boldsymbol{e}_\rho \equiv [\boldsymbol{e}_\mu, \boldsymbol{e}_\nu] = \nabla_\mu \boldsymbol{e}_\nu - \nabla_\nu \boldsymbol{e}_\mu = (\Gamma^{\rho}{}_{\nu\mu} - \Gamma^{\rho}{}_{\mu\nu})\boldsymbol{e}_\rho = 2\Gamma^{\rho}{}_{[\nu\mu]}\boldsymbol{e}_\rho;$$

i.e.,

$$\Gamma^{\rho}{}_{[\mu\nu]} = -\frac{1}{2}c_{\mu\nu}{}^{\rho}; \qquad \Gamma_{\rho[\mu\nu]} = -\frac{1}{2}c_{\mu\nu\rho}. \tag{8.34}$$

(4) This, combined with step (2) yields the desired formula for $\Gamma_{\mu\beta\gamma}$.]

§8.6. LOCAL LORENTZ FRAMES: MATHEMATICAL DISCUSSION

An observer falling freely in curved spacetime makes measurements in his local Lorentz frame. What he discovers has been discussed extensively in Parts I and II of this book. Try now to derive his basic discoveries from the formalism of the last section.

Local Lorentz frame:

Pick an event $\mathscr{P}_0$ on the observer's world line. His local Lorentz frame there is a coordinate system $x^\alpha(\mathscr{P})$ in which

$$g_{\alpha\beta} \equiv \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta \equiv \frac{\partial \mathscr{P}}{\partial x^\alpha} \cdot \frac{\partial \mathscr{P}}{\partial x^\beta} = \eta_{\alpha\beta} \text{ at } \mathscr{P}_0 \tag{8.35a}$$

(Lorentz metric at $\mathscr{P}_0$), and in which

$$\partial g_{\alpha\beta}/\partial x^\mu = 0 \text{ at } \mathscr{P}_0 \tag{8.35b}$$

(metric as Lorentz as possible near $\mathscr{P}_0$). [See equation (8.15).] In addition, by virtue of equations (8.24),

$$\Gamma^\alpha{}_{\beta\gamma} = 0 \text{ at } \mathscr{P}_0 \tag{8.36}$$

(no "correction terms" in covariant derivatives). Of course, the observer must be at rest in his local Lorentz frame; i.e., his world line must be

$$x^j = x^j(\mathscr{P}_0) = \text{constant}; \qquad x^0 \text{ varying.} \tag{8.37}$$

Origin falls freely along a geodesic

Query: Equations (8.35) to (8.37) guarantee that the observer is at rest in a local Lorentz frame. Do they imply that he is freely falling? (They should!) *Answer:* Calculate the observer's 4-acceleration $\boldsymbol{a} = d\boldsymbol{u}/d\tau$ (notation of chapter 6) $= \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u}$ (notation of this chapter). His 4-velocity, calculated from equation (8.37) is

$$\boldsymbol{u} = (dx^\alpha/d\tau)\boldsymbol{e}_\alpha = (dx^0/d\tau)\boldsymbol{e}_0 \underset{\uparrow}{=} \boldsymbol{e}_0; \tag{8.38}$$
$$\begin{bmatrix}\text{because } \boldsymbol{u} \text{ and } \boldsymbol{e}_0 \text{ both} \\ \text{have unit length}\end{bmatrix}$$

so his 4-acceleration is

$$\begin{aligned}\boldsymbol{a} = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = \boldsymbol{\nabla}_0\boldsymbol{e}_0 &= \Gamma^\alpha{}_{00}\boldsymbol{e}_\alpha \\ &= 0 \text{ at } \mathscr{P}_0.\end{aligned} \tag{8.39}$$

Thus, he is indeed freely falling ($\boldsymbol{a} = 0$); and he moves along a geodesic ($\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0$).

Freely falling particles move on straight lines

Query: Do freely falling particles move along straight lines ($d^2x^\alpha/d\tau^2 = 0$) in the observer's local Lorentz frame at $\mathscr{P}_0$? (They should!) *Answer:* A freely falling particle experiences zero 4-acceleration

$$\boldsymbol{a}_{\text{particle}} = \boldsymbol{\nabla}_{\boldsymbol{u}_{\text{particle}}}\boldsymbol{u}_{\text{particle}} = 0;$$

i.e., it parallel-transports its 4-velocity; i.e., it moves along a geodesic of spacetime with affine parameter equal to its proper time. The geodesic equation for its world line, in local Lorentz coordinates, says

$$\frac{d^2x^\alpha}{d\tau^2} = -\Gamma^\alpha{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}$$
$$= 0 \text{ at } \mathcal{P}_o.$$

The particle's world line is, indeed, straight at $\mathcal{P}_o$.

Basis vectors at origin are Fermi-Walker transported

Query: Does the freely falling observer Fermi-Walker-transport his spatial basis vectors $\boldsymbol{e}_j$; i.e., can he attach them to gyroscopes that he carries? (He should be able to!) *Answer:* Fermi-Walker transport (Box 6.2) would say

$$\underbrace{\frac{d\boldsymbol{e}_j}{d\tau}}_{\text{old notation}} \equiv \underbrace{\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{e}_j}_{\text{new notation}} = \boldsymbol{u}(\boldsymbol{a}\cdot\boldsymbol{e}_j) - \boldsymbol{a}(\boldsymbol{u}\cdot\boldsymbol{e}_j).$$

But $\boldsymbol{u} = \boldsymbol{e}_0$, $\boldsymbol{e}_0\cdot\boldsymbol{e}_j = 0$, and $\boldsymbol{a} = 0$ for the observer; so Fermi-Walker transport in this case reduces to parallel transport along $\boldsymbol{e}_0$: thus $\boldsymbol{\nabla}_0\boldsymbol{e}_j = 0$. This is, indeed, how $\boldsymbol{e}_j$ is transported through $\mathcal{P}_o$, because

$$\boldsymbol{\nabla}_0\boldsymbol{e}_j = \Gamma^\alpha{}_{j0}\boldsymbol{e}_\alpha = 0 \text{ at } \mathcal{P}_o.$$

§8.7. GEODESIC DEVIATION AND THE RIEMANN CURVATURE TENSOR

"Gravitation is a manifestation of spacetime curvature, and that curvature shows up in the deviation of one geodesic from a nearby geodesic (relative acceleration of test particles)." To make this statement precise, first quantify the "deviation" or "relative acceleration" of neighboring geodesics.

Focus attention on a family of geodesics $\mathcal{P}(\lambda, n)$; see Figure 8.4. The smoothly varying parameter n ("selector parameter") distinguishes one geodesic from the next. For fixed n, $\mathcal{P}(\lambda, n)$ is a geodesic with affine parameter λ and with tangent vector

$$\boldsymbol{u} = \partial\mathcal{P}/\partial\lambda; \tag{8.40}$$

thus $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0$ (geodesic equation). The vector

$$\boldsymbol{n} \equiv \partial\mathcal{P}/\partial n \tag{8.41}$$

measures the separation between points with the same value of λ on neighboring geodesics.

An observer falling freely along the "fiducial geodesic" $n = 0$ watches a test particle fall along the "test geodesic" $n = 1$. The velocity of the test particle relative

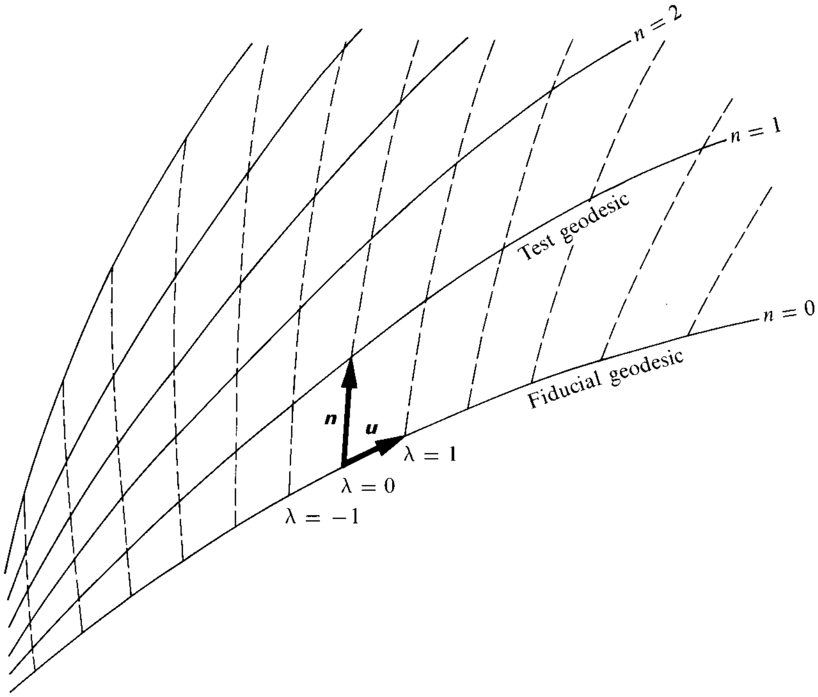


Figure 8.4.
A family of geodesics $\mathscr{P}(\lambda, n)$. The selector parameter n tells "which" geodesic; the affine parameter λ tells "where" on a given geodesic. The separation vector $\boldsymbol{n} \equiv \partial\mathscr{P}/\partial n$ at a point $\mathscr{P}(\lambda, 0)$ along the fiducial geodesic, $n = 0$, reaches (approximately) to the point $\mathscr{P}(\lambda, 1)$ with the same value of λ on the test geodesic, $n = 1$.

to him he quantifies by $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{n}$. This relative velocity, like the separation vector $\boldsymbol{n}$, is an arbitrary "initial condition." Not arbitrary, however, is the "relative acceleration," $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{n}$ of the test particle relative to the observer (see Boxes 11.2 and 11.3). It would be zero in flat spacetime. In curved spacetime, it is given by

Riemann curvature tensor defined by relative acceleration of geodesics

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{n} + \textbf{\textit{Riemann}}\,(\ldots, \boldsymbol{u}, \boldsymbol{n}, \boldsymbol{u}) = 0, \tag{8.42}$$

or, in component notation,

$$\frac{D^2 n^\alpha}{d\lambda^2} + R^\alpha{}_{\beta\gamma\delta}u^\beta n^\gamma u^\delta = 0. \tag{8.43}$$

This equation serves as a definition of the *"Riemann curvature tensor;"* and it can also be used to derive the following expressions for the components of ***Riemann*** *in a coordinate basis:*

Components of ***Riemann***

$$\begin{aligned} R^\alpha{}_{\beta\gamma\delta} &= \langle \boldsymbol{d}x^\alpha, [\boldsymbol{\nabla}_\gamma, \boldsymbol{\nabla}_\delta]\boldsymbol{e}_\beta\rangle \\ &= \frac{\partial\Gamma^\alpha{}_{\beta\delta}}{\partial x^\gamma} - \frac{\partial\Gamma^\alpha{}_{\beta\gamma}}{\partial x^\delta} + \Gamma^\alpha{}_{\mu\gamma}\Gamma^\mu{}_{\beta\delta} - \Gamma^\alpha{}_{\mu\delta}\Gamma^\mu{}_{\beta\gamma}. \end{aligned} \tag{8.44}$$

(For proof, read Box 11.4, Box 11.5, and exercise 11.3, in that order.) For a glimpse of the man who first analyzed the curvature of spaces with three and more dimensions, see Box 8.5.

Effects of curvature

Spacetime curvature causes not only geodesic deviation, but also route dependence in parallel transport (parallel transport around a closed curve changes a vector or tensor—Box 11.7); it causes covariant derivatives to fail to commute [equation (8.44)]; and it prevents the existence of a global Lorentz coordinate system (§11.5).

At first sight one might think ***Riemann*** has $4 \times 4 \times 4 \times 4 = 256$ independent components. But closer examination (§13.5) reveals a variety of symmetries

Symmetries of ***Riemann***

$$R_{\alpha\beta\gamma\delta} = R_{[\alpha\beta][\gamma\delta]} = R_{[\gamma\delta][\alpha\beta]}, \qquad R_{[\alpha\beta\gamma\delta]} = 0, \qquad R_{\alpha[\beta\gamma\delta]} = 0 \tag{8.45}$$

Box 8.5 GEORG FRIEDRICH BERNHARD RIEMANN

September 17, 1826, Breselenz, Hanover—July 20, 1866, Selasca, Lake Maggiore

With his famous doctoral thesis of 1851, "Foundations for a general theory of functions of a single complex variable," Riemann founded one branch of modern mathematics (the theory of Riemann surfaces); and with his famous lecture of three years later founded another (Riemannian geometry). These and other writings will be found in his collected works, edited by H. Weber (1953).

"The properties which distinguish space from other conceivable triply-extended magnitudes are only to be deduced from experience. . . . At every point the three-directional measure of curvature can have an arbitrary value if only the effective curvature of every measurable region of space does not differ noticeably from zero." [G. F. B. Riemann, "On the hypotheses that lie at the foundations of geometry," *Habilitationsvorlesung* of June 10, 1854, on entry into the philosophical faculty of the University of Göttingen.]

Dying of tuberculosis twelve years later, occu-

(antisymmetry on first two indices; antisymmetry on last two; symmetry under exchange of first pair with last pair; vanishing of completely antisymmetric parts). These reduce ***Riemann*** (in four dimensions) from 256 to 20 independent components.

Besides these algebraic symmetries, ***Riemann*** possesses differential symmetries called *"Bianchi identities,"*

$$R^{\alpha}{}_{\beta[\lambda\mu;\nu]} = 0, \tag{8.46}$$

Bianchi identities

which have deep geometric significance (Chapter 15).

From ***Riemann*** one can form several other curvature tensors by contraction. The easiest to form are the *"Ricci curvature tensor,"*

pied with an attempt at a unified explanation of gravity and electromagnetism, Riemann communicated to Betti his system of characterization of multiply-connected topologies (which opened the door to the view of electric charge as "lines of force trapped in the topology of space"), making use of numbers that today are named after Betti but that are identified with a symbol, R_n, that honors Riemann.

"A more detailed scrutiny of a surface might disclose that what we had considered an elementary piece in reality has tiny handles attached to it which change the connectivity character of the piece, and that a microscope of ever greater magnification would reveal ever new topological complications of this type, *ad infinitum.* The Riemann point of view allows, also for real space, topological conditions entirely different from those realized by Euclidean space. I believe that only on the basis of the freer and more general conception of geometry which had been brought out by the development of mathematics during the last century, and with an open mind for the imaginative possibilities which it has revealed, can a philosophically fruitful attack upon the space problem be undertaken." H. Weyl (1949, p. 91).

"But . . . physicists were still far removed from such a way of thinking; space was still, for them, a rigid, homogeneous something, susceptible of no change or conditions. Only the genius of Riemann, solitary and uncomprehended, had already won its way by the middle of the last century to a new conception of space, in which space was deprived of its rigidity, and in which its power to take part in physical events was recognized as possible." A. Einstein (1934, p. 68).

Riemann formulated the first known model for superspace (for which see Chapter 43), a superspace built, however, not of the totality of all 3-geometries with positive definite Riemannian metric (the dynamic arena of Einstein's general relativity), but of all conformally equivalent closed Riemannian 2-geometries of the same topology, a type of superspace known today as Teichmüller space, for more on Riemann's contributions to which and the subsequent development of which, see the chapters by L. Bers and J. A. Wheeler in Gilbert and Newton (1970).

Ricci curvature tensor

$$R_{\mu\nu} \equiv R^{\alpha}{}_{\mu\alpha\nu} \underset{\text{[in a coordinate frame]}}{=} \Gamma^{\alpha}{}_{\mu\nu,\alpha} - \Gamma^{\alpha}{}_{\mu\alpha,\nu} + \Gamma^{\alpha}{}_{\beta\alpha}\Gamma^{\beta}{}_{\mu\nu} - \Gamma^{\alpha}{}_{\beta\nu}\Gamma^{\beta}{}_{\mu\alpha}, \tag{8.47}$$

and the *"scalar curvature,"*

Scalar curvature

$$R \equiv R^{\mu}{}_{\mu}. \tag{8.48}$$

But of much greater geometric significance is the *"Einstein curvature tensor"*

Einstein curvature tensor

$$G^{\mu}{}_{\nu} \equiv \frac{1}{2}\varepsilon^{\mu\alpha\beta\gamma}R_{\beta\gamma}{}^{\rho\sigma}\frac{1}{2}\varepsilon_{\nu\alpha\rho\sigma} = R^{\mu}{}_{\nu} - \frac{1}{2}\delta^{\mu}{}_{\nu}R. \tag{8.49}$$

Of all second-rank curvature tensors one can form by contracting ***Riemann***, only ***Einstein*** = ***G*** retains part of the Bianchi identities (8.46): it satisfies

Contracted Bianchi identities

$$G^{\mu\nu}{}_{;\nu} = 0. \tag{8.50}$$

For the beautiful geometric meaning of these *"contracted Bianchi identities"* ("the boundary of a boundary is zero"), see Chapter 15.

Box 8.6 summarizes the above equations describing curvature, as well as the fundamental equations for covariant derivatives.

EXERCISE

[The following exercises from Track 2 are appropriate for the Track-1 reader who wishes to solidfy his understanding of curvature: 11.6, 11.9, 11.10, 13.7–11, and 14.3.]

Exercise 8.16. SOME USEFUL FORMULAS IN COORDINATE FRAMES

In any coordinate frame, define g to be the determinant of the matrix $g_{\alpha\beta}$ [equation 8.11]. Derive the following relations, valid in any coordinate frame.

(a) Contraction of connection coefficients:

$$\Gamma^{\alpha}{}_{\beta\alpha} = (\ln\sqrt{-g})_{,\beta}. \tag{8.51a}$$

[*Hint:* Use the results of exercise 5.5.]

(b) Components of Ricci tensor:

$$R_{\alpha\beta} = \frac{1}{\sqrt{-g}}(\sqrt{-g}\,\Gamma^{\mu}{}_{\alpha\beta})_{,\mu} - (\ln\sqrt{-g})_{,\alpha\beta} - \Gamma^{\mu}{}_{\nu\alpha}\Gamma^{\nu}{}_{\beta\mu}. \tag{8.51b}$$

(c) Divergence of a vector A^{α} or *antisymmetric* tensor $F^{\alpha\beta}$:

$$A^{\alpha}{}_{;\alpha} = \frac{1}{\sqrt{-g}}(\sqrt{-g}\,A^{\alpha})_{,\alpha}, \qquad F^{\alpha\beta}{}_{;\beta} = \frac{1}{\sqrt{-g}}(\sqrt{-g}\,F^{\alpha\beta})_{,\beta}. \tag{8.51c}$$

(d) Integral of a scalar field Ψ over the proper volume of a 4-dimensional region $\mathcal{V}$:

$$\int_{\mathcal{V}} \Psi\, d(\text{proper volume}) = \int_{\mathcal{V}} \Psi\sqrt{-g}\, dt\, dx\, dy\, dz. \tag{8.51d}$$

[*Hint:* In a local Lorentz frame, d(proper volume) $= d\hat{t}\, d\hat{x}\, d\hat{y}\, d\hat{z}$. Use a Jacobian to transform this volume element to the given coordinate frame, and prove from the transformation law

$$g_{\alpha\beta} = \frac{\partial x^{\hat{\mu}}}{\partial x^{\alpha}} \frac{\partial x^{\hat{\nu}}}{\partial x^{\beta}} \eta_{\mu\nu}$$

that the Jacobian is equal to $\sqrt{-g}$.]

Box 8.6 COVARIANT DERIVATIVE AND CURVATURE: FUNDAMENTAL EQUATIONS

Entity	*Abstract notation*	*Component notation*
Covariant Derivative	$\nabla_{\boldsymbol{u}} \boldsymbol{T} = \nabla \boldsymbol{T}(\ldots,\ldots,\boldsymbol{u})$	$T^{\beta}{}_{\alpha;\gamma}u^{\gamma} = DT^{\beta}{}_{\alpha}/d\lambda \qquad (\boldsymbol{u} = d\mathscr{P}/d\lambda)$ $= \dfrac{dT^{\beta}{}_{\alpha}}{d\lambda} + (\Gamma^{\beta}{}_{\nu\mu}T^{\nu}{}_{\alpha} - \Gamma^{\nu}{}_{\alpha\mu}T^{\beta}{}_{\nu})u^{\mu}$
	$\nabla_{\boldsymbol{u}} f = \partial_{\boldsymbol{u}} f = \langle \boldsymbol{d}f, \boldsymbol{u}\rangle$	$f_{;\alpha}u^{\alpha} = f_{,\alpha}u^{\alpha}$
algebraic properties (Exercise 8.11)	$\nabla_{a\boldsymbol{u}+b\boldsymbol{v}}\boldsymbol{T} = a\nabla_{\boldsymbol{u}}\boldsymbol{T} + b\nabla_{\boldsymbol{v}}\boldsymbol{T}$ $\nabla_{\boldsymbol{u}}(\boldsymbol{S} + \boldsymbol{M}) = \nabla_{\boldsymbol{u}}\boldsymbol{S} + \nabla_{\boldsymbol{u}}\boldsymbol{M}$ $\nabla_{\boldsymbol{u}}\boldsymbol{w} - \nabla_{\boldsymbol{w}}\boldsymbol{u} = [\boldsymbol{u}, \boldsymbol{w}]$ for $\boldsymbol{u}, \boldsymbol{w}$ both vector fields	$T^{\beta}{}_{\alpha;\gamma}(au^{\gamma} + bv^{\gamma}) = aT^{\beta}{}_{\alpha;\gamma}u^{\gamma} + bT^{\beta}{}_{\alpha;\gamma}v^{\gamma}$ $(S^{\beta}{}_{\alpha} + M^{\beta}{}_{\alpha})_{;\gamma}u^{\gamma} = S^{\beta}{}_{\alpha;\gamma}u^{\gamma} + M^{\beta}{}_{\alpha;\gamma}u^{\gamma}$ $\Gamma^{\rho}{}_{[\mu\nu]} = -\frac{1}{2}c_{\mu\nu}{}^{\rho}$ [equation (8.34)]
chain rule	$\nabla_{\boldsymbol{u}}(\boldsymbol{A} \otimes \boldsymbol{B}) = (\nabla_{\boldsymbol{u}}\boldsymbol{A}) \otimes \boldsymbol{B} + \boldsymbol{A} \otimes (\nabla_{\boldsymbol{u}}\boldsymbol{B})$ $\nabla_{\boldsymbol{u}}(f\boldsymbol{A}) = (\nabla_{\boldsymbol{u}}f)\boldsymbol{A} + f\nabla_{\boldsymbol{u}}\boldsymbol{A}$	$(A^{\alpha}{}_{\beta}B_{\gamma})_{;\mu}u^{\mu} = A^{\alpha}{}_{\beta;\mu}B_{\gamma}u^{\mu} + A^{\alpha}{}_{\beta}B_{\gamma;\mu}u^{\mu}$ $(fA^{\alpha}{}_{\beta})_{;\mu}u^{\mu} = f_{,\mu}A^{\alpha}{}_{\beta}u^{\mu} + fA^{\alpha}{}_{\beta;\mu}u^{\mu}$
$\nabla_{\boldsymbol{u}}$ and contraction commute	$\nabla_{\boldsymbol{u}}$(contraction of $\boldsymbol{S}$) = (contraction of $\nabla_{\boldsymbol{u}}\boldsymbol{S}$)	$\left(\sum_{\alpha} S^{\alpha}{}_{\alpha\gamma}\right)_{;\mu} u^{\mu} = \sum_{\alpha} (S^{\alpha}{}_{\alpha\gamma;\mu}u^{\mu})$
*metric covariantly constant	$\nabla_{\boldsymbol{u}}\boldsymbol{g} = 0$	$g_{\alpha\beta;\gamma}u^{\gamma} = 0$
Gradient	$\nabla \boldsymbol{T}$	$T^{\beta}{}_{\alpha;\gamma} = T^{\beta}{}_{\alpha,\gamma} + \Gamma^{\beta}{}_{\mu\gamma}T^{\mu}{}_{\alpha} - \Gamma^{\mu}{}_{\alpha\gamma}T^{\beta}{}_{\mu}$
Connection Coefficients	$\Gamma^{\alpha}{}_{\beta\gamma} = \langle \boldsymbol{\omega}^{\alpha}, \nabla_{\gamma}\boldsymbol{e}_{\beta}\rangle$ $= -\langle \nabla_{\gamma}\boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta}\rangle$	$\Gamma^{\alpha}{}_{\beta\gamma} = g^{\alpha\mu}\Gamma_{\mu\beta\gamma}$,* $\Gamma_{\mu\beta\gamma} = \frac{1}{2}(g_{\mu\beta,\gamma} + g_{\mu\gamma,\beta} - g_{\beta\gamma,\mu}$ $+ c_{\mu\beta\gamma} + c_{\mu\gamma\beta} - c_{\beta\gamma\mu})$* $c_{\beta\gamma\mu} = g_{\mu\alpha}c_{\beta\gamma}{}^{\alpha}$ $= g_{\mu\alpha}\langle \boldsymbol{\omega}^{\alpha}, [\boldsymbol{e}_{\beta}, \boldsymbol{e}_{\gamma}]\rangle$*
*Local Lorentz frame at $\mathscr{P}_0$		Coordinate system with $g_{\mu\nu}(\mathscr{P}_0) = \eta_{\mu\nu}, \qquad \Gamma^{\alpha}{}_{\beta\gamma}(\mathscr{P}_0) = 0$
Parallel transport	$\nabla_{\boldsymbol{u}}\boldsymbol{S} = 0$	$S^{\alpha}{}_{\beta;\gamma}u^{\gamma} = 0$

Box 8.6 (*continued*)

Entity	*Abstract notation*	*Component notation*
Geodesic Equation	$\nabla_{\boldsymbol{u}} \boldsymbol{u} = 0$	$d^2x^\alpha/d\lambda^2 + \Gamma^\alpha{}_{\mu\nu}(dx^\mu/d\lambda)(dx^\nu/d\lambda) = 0$ in a coordinate basis
Riemann Curvature Tensor	***Riemann*** $(\boldsymbol{\sigma}, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) \equiv \langle \boldsymbol{\sigma}, \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}\rangle$ $\mathcal{R}(\boldsymbol{A}, \boldsymbol{B}) \equiv [\nabla_{\boldsymbol{A}}, \nabla_{\boldsymbol{B}}] - \nabla_{[\boldsymbol{A},\boldsymbol{B}]}$ (not track-one formulas; see Chapter 11)	$R^\alpha{}_{\beta\gamma\delta} = \dfrac{\partial\Gamma^\alpha{}_{\beta\delta}}{\partial x^\gamma} - \dfrac{\partial\Gamma^\alpha{}_{\beta\gamma}}{\partial x^\delta} + \Gamma^\alpha{}_{\mu\gamma}\Gamma^\mu{}_{\beta\delta} - \Gamma^\alpha{}_{\mu\delta}\Gamma^\mu{}_{\beta\gamma}$ in coordinate frame [see equation (11.13) for formula in non-coordinate frame]
Ricci Curvature Tensor	***Ricci*** = contraction on slots 1 and 3 of ***Riemann***	$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu} = \Gamma^\alpha{}_{\mu\nu,\alpha} - \Gamma^\alpha{}_{\mu\alpha,\nu} + \Gamma^\alpha{}_{\beta\alpha}\Gamma^\beta{}_{\mu\nu} - \Gamma^\alpha{}_{\beta\nu}\Gamma^\beta{}_{\mu\alpha}$ in coordinate frame
*Curvature Scalar	R = (contraction of ***Ricci***)	$R = R^\alpha{}_\alpha$
*Einstein Curvature Tensor	$\boldsymbol{G} = \boldsymbol{Ricci} - \frac{1}{2}\boldsymbol{g}R$	$G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R$ Useful formulas for computing $G^\alpha{}_\beta$ (derived in §14.2): $G^0{}_0 = -(R^{12}{}_{12} + R^{23}{}_{23} + R^{31}{}_{31})$, $G^0{}_1 = R^{02}{}_{12} + R^{03}{}_{13}$, $G^1{}_1 = -(R^{02}{}_{02} + R^{03}{}_{03} + R^{23}{}_{23})$, $G^1{}_2 = R^{10}{}_{20} + R^{13}{}_{23}$, etc.
*Symmetries of Curvature Tensors		$R_{\alpha\beta\gamma\delta} = R_{[\alpha\beta][\gamma\delta]} = R_{[\gamma\delta][\alpha\beta]}$, $R_{[\alpha\beta\gamma\delta]} = 0$, $R_{\alpha[\beta\gamma\delta]} = 0$ $R_{\alpha\beta} = R_{(\alpha\beta)}$, $G_{\alpha\beta} = G_{(\alpha\beta)}$
Bianchi Identities		$R^\alpha{}_{\beta[\mu\nu;\lambda]} = 0$
*Contracted Bianchi Identities		$G^{\alpha\beta}{}_{;\beta} = 0$
Geodesic Deviation	$\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n}$ + ***Riemann*** $(\ldots, \boldsymbol{u}, \boldsymbol{n}, \boldsymbol{u}) = 0$	$\dfrac{D^2 n^\alpha}{d\lambda^2} + R^\alpha{}_{\beta\gamma\delta}u^\beta n^\gamma u^\delta = 0$
Parallel Transport around closed curve (§11.4)	$\delta\boldsymbol{A}$ + ***Riemann*** $(\ldots\boldsymbol{A}, \boldsymbol{u}, \boldsymbol{v}) = 0$ if $\boldsymbol{u}$, $\boldsymbol{v}$ are edges of curve	$\delta A^\alpha + R^\alpha{}_{\beta\gamma\delta}A^\beta u^\gamma v^\delta = 0$

*If metric is absent, these starred formulas cannot be formulated. All other formulas are valid in absence of metric.

CHAPTER **9**

DIFFERENTIAL TOPOLOGY

In analytic geometry, many relations which are independent of any frame must be expressed with respect to some particular frame. It is therefore preferable to devise new methods—methods which lead directly to intrinsic properties without any mention of coordinates. The development of the topology of general spaces and of the objects which occur in them, as well as the development of the geometry of general metric spaces, are steps in this direction.

KARL MENGER, in Schilpp (1949), p. 467.

§9.1. GEOMETRIC OBJECTS IN METRIC-FREE, GEODESIC-FREE SPACETIME

This chapter is entirely Track 2.
It depends on no preceding Track-2 material.
It is needed as preparation for
(1) Chapters 10–13 (differential geometry; Newtonian gravity), and
(2) Box 30.1 (mixmaster cosmology).
It will be helpful in
(1) Chapter 14 (calculation of curvature) and in
(2) Chapter 15 (Bianchi identities).

Curved spacetime without metric or geodesics or parallel transport, i.e., "differential topology," is the subject of this easy chapter. It is easy because all the necessary geometric objects (event, curve, vector, 1-form, tensor) are already familiar from flat spacetime. Yet it is also necessary, because one's viewpoint must be refined when one abandons the Lorentz metric of flat spacetime.

Events

The primitive concept of an event $\mathscr{P}$ (Figure 1.2) needs no refinement. The essential property here is identifiability, which is not dependent on the Lorentz metric structure of spacetime.

Metric is abandoned

Curves

Geometric concepts must be refined

Again no refinement. A "curve" $\mathcal{P}(\lambda)$ is also too primitive to care whether spacetime has a metric—*except* that, with metric gone, there is no concept of "proper length" along the curve. This is in accord with Newton's theory of gravity, where one talks of the lengths of curves in "space," but never in "spacetime."

Vectors

Here refinement is needed. In special relativity one could dress primitive ("identifiable") events in enough algebraic plumage to talk of vectors as differences $\mathcal{P} - \mathcal{Q}$ between "algebraic" events. Now the plumage is gone, and the old bilocal ("point for head and point for tail") version of a vector must be replaced by a purely local version (§9.2). Also vectors cannot be moved around; each vector must be attached to a specific event (§§9.2 and 9.3).

1-Forms

Almost no refinement needed, except that, with metric gone, there is no way to tell which 1-form corresponds to a given vector (no way to raise and lower indices), and each 1-form must be attached to a specific event (§9.4).

Tensors

Again almost no refinement, except that each slot of a tensor is specific: if it accepts vectors, then it cannot accommodate 1-forms, and conversely (no raising and lowering of indices); also, each tensor must be attached to a specific event (§9.5).

§9.2. "VECTOR" AND "DIRECTIONAL DERIVATIVE" REFINED INTO TANGENT VECTOR

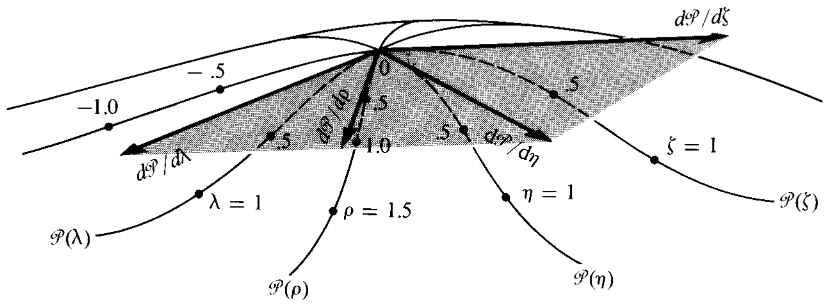
Flat spacetime can accommodate several equivalent definitions of a vector (§2.3): a vector is an arrow reaching from an event $\mathcal{P}_0$ to an event $\mathcal{Q}_0$; it is the parameterized straight line, $\mathcal{P}(\lambda) = \mathcal{P}_0 + \lambda(\mathcal{Q}_0 - \mathcal{P}_0)$ extending from $\mathcal{P}_0$ at $\lambda = 0$ to $\mathcal{Q}_0$ at $\lambda = 1$; it is the rate of change of the point $\mathcal{P}(\lambda)$ with increasing λ, $d\mathcal{P}/d\lambda$.

Old definitions of vector break down when metric is abandoned

With Lorentz metric gone, the "arrow" definition and the "parametrized-straight line" definition must break down. By what route is the arrow or line to be laid out between $\mathcal{P}_0$ and $\mathcal{Q}_0$? There is no concept of straightness; all routes are equally straight or bent.

Such fuzziness forces one to focus on the "rate-of-change-of-point-along-curve"

Box 9.1 TANGENT VECTORS AND TANGENT SPACE



A tangent vector $d\mathscr{P}/d\lambda$ is defined to be "the limit, when $N \longrightarrow \infty$, of N times the displacement of $\mathscr{P}$ as λ ranges from 0 to $1/N$." One cannot think of this final displacement $d\mathscr{P}/d\lambda$ as lying in spacetime; fuzziness forbids (no concept of straightness). Instead, one visualizes $d\mathscr{P}/d\lambda$ as lying in a "tangent plane" or "tangent space," which makes contact with spacetime only at $\mathscr{P}(0)$, the event where $d\mathscr{P}/d\lambda$ is evaluated. All other tangent vectors at $\mathscr{P}(0)$—e.g., $d\mathscr{P}/d\rho$, $d\mathscr{P}/d\eta$, $d\mathscr{P}/d\zeta$—lie in this same tangent space.

To make precise these concepts of tangent space and tangent vector, one may regard spacetime as embedded in a flat space of more than four dimensions. One can then perform the limiting process that leads to $d\mathscr{P}/d\lambda$, using straight arrows in the flat embedding space. The result is a higher-dimensional analog of the figure shown above.

But such a treatment is dangerous. It suggests, falsely, that the tangent vector $d\mathscr{P}/d\lambda$ and the tangent space at $\mathscr{P}_0$ depend on how the embedding is done, or depend for their existence on the embedding process. They do not. And to make clear that they do not is one motivation for defining the directional derivative operator "$d/d\lambda$" to be the tangent vector, rather than using Cartan's more pictorial concept "$d\mathscr{P}/d\lambda$".

definition, $d\mathscr{P}/d\lambda$. It, under the new name *"tangent vector,"* is explored briefly in Box 9.1, and in greater depth in the following paragraphs.

Even "$d\mathscr{P}/d\lambda$" is a fuzzy definition of tangent vector, most mathematicians would argue. More acceptable, they suggest, is this definition: *the tangent vector* $\boldsymbol{u}$ *to a curve* $\mathscr{P}(\lambda)$ *is the directional derivative operator along that curve*

Best new definition: "tangent vector equals directional derivative operator" $\boldsymbol{u} = d/d\lambda$

$$\boldsymbol{u} = \partial_{\boldsymbol{u}} = (d/d\lambda)_{\text{along curve}}. \tag{9.1}$$

Tangent vector equals directional derivative operator? Preposterous! A vector started out as a happy, irresponsible trip from $\mathscr{P}_0$ to $\mathscr{Q}_0$. It ended up loaded with the social responsibility to tell how something else changes at $\mathscr{P}_0$. At what point did the vector get saddled with this unexpected load? And did it really change its character all that much, as it seems to have done? For an answer, go back and try to redo the "rate-of-change-of-point" definition, $d\mathscr{P}/d\lambda$, in the form of a limiting process:

0. Choose a curve $\mathscr{P}(\lambda)$ whose tangent vector $d\mathscr{P}/d\lambda$ at $\lambda = 0$ is desired.
1. Take the displacement of $\mathscr{P}$ as λ ranges from 0 to 1; that is *not* $d\mathscr{P}/d\lambda$.
2. Take twice the displacement of $\mathscr{P}$ as λ ranges from 0 to $\frac{1}{2}$; that is *not* $d\mathscr{P}/d\lambda$.

N. Take N times the displacement of $\mathscr{P}$ as λ ranges from 0 to $1/N$; that is *not* $d\mathscr{P}/d\lambda$.

∞. Take the limit of such displacements as $N \longrightarrow \infty$; that *is* $d\mathscr{P}/d\lambda$.

This definition has the virtue that $d\mathscr{P}/d\lambda$ describes the properties of the curve $\mathscr{P}(\lambda)$, not over the huge range from $\lambda = 0$ to $\lambda = 1$, where the curve might be doing wild things, but only in an infinitesimal neighborhood of the point $\mathscr{P}_0 = \mathscr{P}(0)$.

Alternative definition, $\boldsymbol{u} = d\mathscr{P}/d\lambda$, requires embedding in flat space of higher dimensionality

The deficiency in this definition is that no meaning is assigned to steps $1, 2, \ldots, N, \ldots$, so there is nothing, yet, to take the limit of. To make each "displacement of $\mathscr{P}$" a definite mathematical object in a space where "limit" has a meaning, one can imagine the original manifold to be a low-dimensional surface in some much higher-dimensional *flat* space. Then $\mathscr{P}(1/N) - \mathscr{P}(0)$ is just a straight arrow connecting two points, i.e. a segment of a straight line, which, in general, will not lie in the surface itself—see Box 9.1. The resulting mental picture of a tangent vector makes its essential properties beautifully clear, but at the cost of some artifacts. The picture relies on a specific but arbitrary way of embedding the manifold of interest (metric-free spacetime) in an extraneous flat space. In using this picture, one must ignore everything that depends on the peculiarities of the embedding. One must think like the chemist, who uses tinkertoy molecular models to visualize many essential properties of a molecule clearly, but easily ignores artifacts of the model (colors of the atoms, diameters of the pegs, its tendency to collapse) that do not mimic quantum-mechanical reality.

Élie Cartan's approach to differential geometry, including the $d\mathscr{P}/d\lambda$ idea of a tangent vector, suggests that he always thought of manifolds as embedded in flat spaces this way, and relied on insights that he did not always formalize to separate the essential geometry of these pictures from their embedding-dependent details. Acceptance of his methods of calculation came late. Mathematicians, who mistrusted their own ability to distinguish fact from artifact, exacted this price for acceptance: stop talking about the movement of the point itself, and start dealing only with concrete measurable changes that take place within the manifold, changes in any or all scalar functions f as the point moves. The limiting process then reads:

Refinement of $d\mathscr{P}/d\lambda$ into $d/d\lambda$

0. Choose a curve $\mathscr{P}(\lambda)$ whose tangent vector at $\lambda = 0$ is desired.
1. Compute the number $f[\mathscr{P}(1)] - f[\mathscr{P}(0)]$, which measures the change in f as the point $\mathscr{P}(\lambda)$ moves from $\mathscr{P}_0 = \mathscr{P}(0)$ to $\mathscr{Q}_0 = \mathscr{P}(1)$.
2. Compute $2\{f[\mathscr{P}(\frac{1}{2})] - f[\mathscr{P}(0)]\}$, which is twice the change in f as the point goes from $\mathscr{P}(0)$ to $\mathscr{P}(\frac{1}{2})$.

N. Compute $N\{f[\mathscr{P}(1/N)] - f[\mathscr{P}(0)]\}$, which is N times the change in f as the point goes from $\mathscr{P}(0)$ to $\mathscr{P}(1/N)$.

∞. Same in the limit as $N \longrightarrow \infty$: (change in f) $= df/d\lambda$.

0. The vector is not itself the change in f. It is instead the operation $d/d\lambda$, which, when applied to f, gives the change $df/d\lambda$. Thus

$$\text{tangent vector} = d/d\lambda$$

[cf. definition (9.1)].

The operation $d/d\lambda$ clearly involves nothing but the last steps $N \longrightarrow \infty$ in this limiting process, and only those aspects of these steps that are independent of f. But this means it involves the infinitesimal displacements of the point $\mathscr{P}$ *and nothing more.*

One who wishes both to stay in touch with the present and to not abandon Cartan's deep geometric insight (Box 9.1) can seek to keep alive a distinction between:

(A) the tangent vector itself in the sense of Cartan, the displacement $d\mathscr{P}/d\lambda$ of a point; and

(B) the "tangent vector operator," or "directional derivative operator," telling what happens to a function in this displacement: (tangent vector operator) $= d/d\lambda$.

However, present practice drops (or, if one will, "slurs") the word "operator" in (B), and uses the phrase "tangent vector" itself for the operator, as will be the practice here from now on. The ideas (A) and (B) should also slur or coalesce in one's mind, so that when one visualizes an embedding diagram with arrows drawn tangent to the surface, one always realizes that the arrow characterizes an infinitesimal motion of a point $d\mathscr{P}/d\lambda$ that takes place purely within the surface, and when one thinks of a derivative operator $d/d\lambda$, one always visualizes this same infinitesimal motion of a point in the manifold, a motion that must occur in constructing any derivative $df(\mathscr{P})/d\lambda$. In this sense, one should regard a vector $d\mathscr{P}/d\lambda \equiv d/d\lambda$ as both "a displacement that carries attention from one point to another" and "a purely geometric object built on points and nothing but points."

The hard-nosed physicist may still be inclined to say "Tangent vector equals directional derivative operator? Preposterous!" Perhaps he will be put at ease by another argument. He is asked to pick an event $\mathscr{P}_0$. At that event he chooses any set of four noncoplanar vectors (vectors defined in whatever way seems reasonable to him); he names them $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$; and he uses them as a basis on which to expand all other vectors at $\mathscr{P}_0$:

$$\boldsymbol{u} = u^\alpha \boldsymbol{e}_\alpha, \qquad \boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha. \tag{9.2}$$

He is asked to construct the four directional derivative operators $\partial_\alpha \equiv \partial_{\boldsymbol{e}_\alpha}$ along his four basis vectors. As in flat spacetime, so also here; the same expansion coefficients that appear in $\boldsymbol{u} = u^\alpha \boldsymbol{e}_\alpha$ also appear in the expansion for the directional derivative:

Isomorphism between directional derivatives and vectors

$$\partial_{\boldsymbol{u}} = u^\alpha\, \partial_\alpha, \qquad \partial_{\boldsymbol{v}} = v^\alpha\, \partial_\alpha. \tag{9.3}$$

Hence, every relation between specific vectors at $\mathscr{P}_0$ induces an identical relation between their differential operators:

$$\begin{aligned} \boldsymbol{u} = a\boldsymbol{w} + b\boldsymbol{v} &\Longleftrightarrow u^\alpha = aw^\alpha + bv^\alpha \\ &\Longleftrightarrow \partial_{\boldsymbol{u}} = a\,\partial_{\boldsymbol{w}} + b\,\partial_{\boldsymbol{v}}. \end{aligned} \tag{9.4}$$

There is a complete "isomorphism" between the vectors and the corresponding directional derivatives. So how can the hard-nosed physicist deny the hard-nosed mathematician the right to identify completely each tangent vector with its directional derivative? No harm is done; no answer to any computation can be affected.

Tangent space defined

This isomorphism extends to the concept *"tangent space."* Because linear relations (such as $\partial_{\boldsymbol{u}} = a\,\partial_{\boldsymbol{w}} + b\,\partial_{\boldsymbol{v}}$) among directional derivatives evaluated *at one and the same point* $\mathscr{P}_0$ are meaningful and obey the usual addition and multiplication rules, these derivative operators form an abstract (but finite-dimensional) vector space called the tangent space at $\mathscr{P}_0$. In an embedding picture (Box 9.1) one uses these derivatives (as operators in the flat embedding space) to construct tangent vectors $\boldsymbol{u} = \partial_{\boldsymbol{u}}\mathscr{P}$, $\boldsymbol{v} = \partial_{\boldsymbol{v}}\mathscr{P}$, in the form of straight arrows. Thereby one identifies the abstract tangent space with the geometrically visualized tangent space.

§9.3. BASES, COMPONENTS, AND TRANSFORMATION LAWS FOR VECTORS

Coordinate-induced basis defined

An especially useful basis in the tangent space at an event $\mathscr{P}_0$ is induced by any coordinate system [four functions, $x^0(\mathscr{P})$, $x^1(\mathscr{P})$, $x^2(\mathscr{P})$, $x^3(\mathscr{P})$]:

$$\boldsymbol{e}_0 \equiv \left(\frac{\partial}{\partial x^0}\right)_{x^1,x^2,x^3} = \begin{pmatrix}\text{directional derivative along the} \\ \text{curve with constant } (x^1, x^2, x^3) \\ \text{and with parameter } \lambda = x^0\end{pmatrix}_{\text{at } \mathscr{P}_0}, \tag{9.5}$$

$$\boldsymbol{e}_1 = \frac{\partial}{\partial x^1}, \qquad \boldsymbol{e}_2 = \frac{\partial}{\partial x^2}, \qquad \boldsymbol{e}_3 = \frac{\partial}{\partial x^3}.$$

(See Figure 9.1.)

Changes of basis: transformation matrices defined

A transformation from one basis to another in the tangent space at $\mathscr{P}_0$, like any change of basis in any vector space, is produced by a nonsingular matrix,

$$\boldsymbol{e}_{\alpha'} = \boldsymbol{e}_\beta L^\beta{}_{\alpha'}; \tag{9.6}$$

and, as always (including the Lorentz frames of flat spacetime), the components of a vector must transform by the inverse matrix

$$u^{\alpha'} = L^{\alpha'}{}_\beta u^\beta; \tag{9.7}$$

$$\|L^{\alpha'}{}_\beta\| = \|L^\beta{}_{\gamma'}\|^{-1}, \text{ i.e., } \begin{cases} L^{\alpha'}{}_\beta L^\beta{}_{\gamma'} = \delta^{\alpha'}{}_{\gamma'}, \\ L^\delta{}_{\alpha'} L^{\alpha'}{}_\beta = \delta^\delta{}_\beta. \end{cases} \tag{9.8}$$

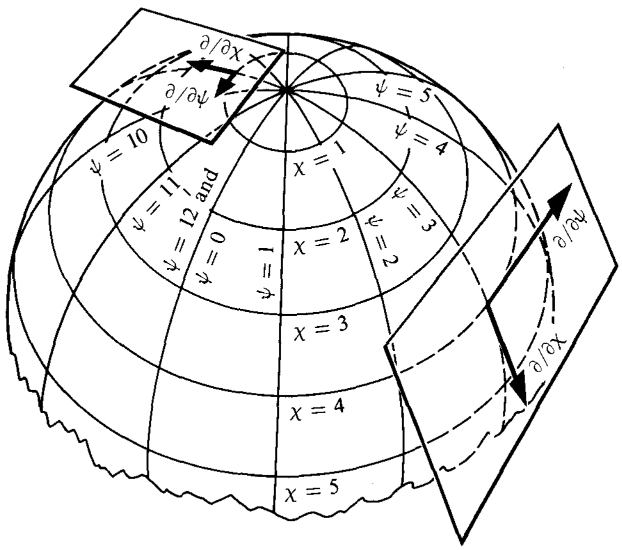


Figure 9.1.
The basis vectors induced, by a coordinate system, into the tangent space at each event. Here a truncated, two-dimensional spacetime is shown (two other dimensions suppressed), with coordinates $\chi(\mathscr{P})$ and $\psi(\mathscr{P})$, and with corresponding basis vectors $\partial/\partial\chi$ and $\partial/\partial\psi$.

This "inverse" transformation law guarantees compatibility between the expansions $\boldsymbol{u} = \boldsymbol{e}_{\alpha'}u^{\alpha'}$ and $\boldsymbol{u} = \boldsymbol{e}_\beta u^\beta$:

$$\begin{aligned}\boldsymbol{u} &= \boldsymbol{e}_{\alpha'}u^{\alpha'} = (\boldsymbol{e}_\gamma L^\gamma{}_{\alpha'})(L^{\alpha'}{}_\beta u^\beta) = \boldsymbol{e}_\gamma \delta^\gamma{}_\beta u^\beta \\ &= \boldsymbol{e}_\beta u^\beta.\end{aligned}$$

In the special case of transformations between coordinate-induced bases, the transformation matrix has a simple form:

$$\frac{\partial}{\partial x^{\alpha'}} = \frac{\partial x^\beta}{\partial x^{\alpha'}}\frac{\partial}{\partial x^\beta} \text{ (by usual rules of calculus),}$$

so

$$L^\beta{}_{\alpha'} = (\partial x^\beta/\partial x^{\alpha'})_{\text{at event } \mathscr{P}_0 \text{ where tangent space lies}}. \tag{9.9}$$

(*Note:* this generalizes the Lorentz-transformation law $x^\beta = \Lambda^\beta{}_{\alpha'}x^{\alpha'}$, which has the differential form $\Lambda^\beta{}_{\alpha'} = \partial x^\beta/\partial x^{\alpha'}$; also, it provides a good way to remember the signs in the Λ matrices.)

§9.4. 1-FORMS

When the Lorentz metric is removed from spacetime, one must sharpen up the concept of a 1-form $\boldsymbol{\sigma}$ by insisting that it, like any tangent vector $\boldsymbol{u}$, be attached to a specific event $\mathscr{P}_0$ in spacetime. The family of surfaces representing $\boldsymbol{\sigma}$ resides in the tangent space at $\mathscr{P}_0$, not in spacetime itself. The piercing of surfaces of $\boldsymbol{\sigma}$ by an arrow $\boldsymbol{u}$ to produce the number $\langle\boldsymbol{\sigma}, \boldsymbol{u}\rangle$ ("bongs of bell") occurs entirely in the tangent space.

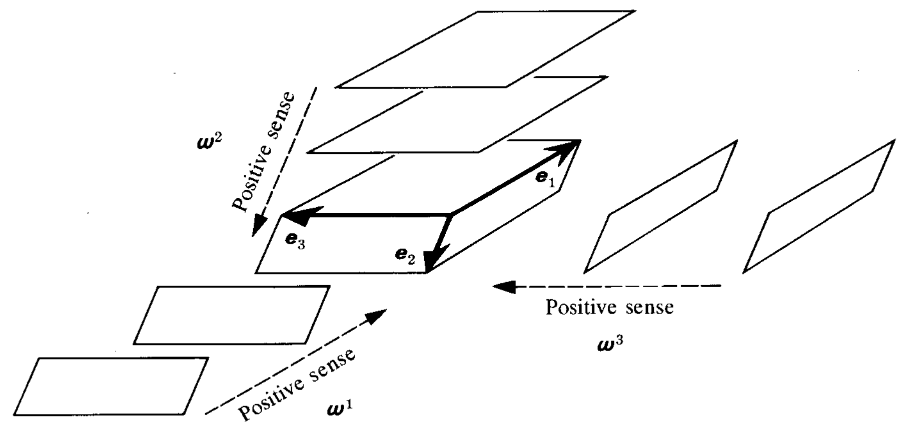


Figure 9.2.
The basis vectors $\boldsymbol{e}_\alpha$ and dual basis 1-forms $\boldsymbol{\omega}^\beta$ in the tangent space of an event $\mathscr{P}_0$. The condition

$$\langle \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \delta^\beta{}_\alpha$$

dictates that the vectors $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ lie parallel to the surfaces of $\boldsymbol{\omega}^1$, and that $\boldsymbol{e}_1$ extend from one surface of $\boldsymbol{\omega}^1$ to the next (precisely 1.00 surfaces pierced).

Notice that this picture could fit perfectly well into a book on X-rays and crystallography. There the vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ would be the edges of a unit cell of the crystal; and the surfaces of $\boldsymbol{\omega}^1$, $\boldsymbol{\omega}^2$, $\boldsymbol{\omega}^3$ would be the surfaces of unit cells. Also, for an X-ray diffraction experiment, with wavelength of radiation and orientation of crystal appropriately adjusted, the successive surfaces of $\boldsymbol{\omega}^1$ would produce Bragg reflection. For other choices of wavelength and orientation, the surfaces of $\boldsymbol{\omega}^2$ or $\boldsymbol{\omega}^3$ would produce Bragg reflection.

Dual basis of 1-forms defined

Given any set of basis vectors $\{\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ at an event $\mathscr{P}_0$, one constructs the *"dual basis"* of 1-forms $\{\boldsymbol{\omega}^0, \boldsymbol{\omega}^1, \boldsymbol{\omega}^2, \boldsymbol{\omega}^3\}$ by choosing the surfaces of $\boldsymbol{\omega}^\beta$ such that

$$\langle \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \delta^\beta{}_\alpha. \tag{9.10}$$

See Figure 9.2. A marvelously simple formalism for calculating and manipulating components of tangent vectors and 1-forms then results:

Component-manipulation formulas

$$\boldsymbol{u} = \boldsymbol{e}_\alpha u^\alpha \quad \text{(definition of components of } \boldsymbol{u}\text{)}, \tag{9.11a}$$

$$\boldsymbol{\sigma} = \sigma_\beta \boldsymbol{\omega}^\beta \quad \text{(definition of components of } \boldsymbol{\sigma}\text{)}, \tag{9.11b}$$

$$u^\alpha = \langle \boldsymbol{\omega}^\alpha, \boldsymbol{u} \rangle \quad \text{(way to calculate components of } \boldsymbol{u}\text{)}, \tag{9.11c}$$

$$\sigma_\beta = \langle \boldsymbol{\sigma}, \boldsymbol{e}_\beta \rangle \quad \text{(way to calculate components of } \boldsymbol{\sigma}\text{)}, \tag{9.11d}$$

$$\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle = \sigma_\alpha u^\alpha \quad \text{(way to calculate } \langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \text{ using components)}, \tag{9.11e}$$

$$\boldsymbol{\omega}^{\alpha'} = L^{\alpha'}{}_\beta \boldsymbol{\omega}^\beta \quad \text{(transformation law for 1-form basis, corresponding to equation 9.6)}, \tag{9.11f}$$

$$\sigma_{\alpha'} = \sigma_\beta L^\beta{}_{\alpha'} \quad \text{(transformation law for 1-form components)}. \tag{9.11g}$$

(Exercise 9.1 below justifies these equations.)

In the absence of a metric, there is no way to pick a specific 1-form $\tilde{\boldsymbol{u}}$ at an event $\mathscr{P}_0$ and say that it corresponds to a specific tangent vector $\boldsymbol{u}$ at $\mathscr{P}_0$. The correspondence set up in flat spacetime,

Correspondence between vectors and 1-forms rubbed out

$$\langle \tilde{\boldsymbol{u}}, \boldsymbol{v} \rangle = \boldsymbol{u} \cdot \boldsymbol{v} \qquad \text{for all } \boldsymbol{v},$$

was rubbed out when "$\cdot$" was rubbed out. Restated in component language: the raising of an index, $u^\alpha = \eta^{\alpha\beta} u_\beta$, is impossible because the $\eta^{\alpha\beta}$ do not exist; similarly, lowering of an index, $u_\beta = \eta_{\beta\alpha} u^\alpha$, is impossible.

Gradient of a function

The 1-form gradient $\boldsymbol{d}f$ was introduced in §2.6 with absolutely no reference to metric. Consequently, it and its mathematical formalism are the same here, without metric, as there with metric, except that, like all other 1-forms, $\boldsymbol{d}f$ now resides in the tangent space rather than in spacetime itself. For example, there is no change in the fundamental equation relating the projection of the gradient to the directional derivative:

$$\langle \boldsymbol{d}f, \boldsymbol{u} \rangle = \underset{\substack{\uparrow \\ \left[\text{old notation for directional derivative}\right]}}{\partial_{\boldsymbol{u}} f} = \underset{\substack{\uparrow \\ \left[\text{new notation; recall } \boldsymbol{u} = \partial_{\boldsymbol{u}}\right]}}{\boldsymbol{u}[f]}. \tag{9.12}$$

Similarly, there are no changes in the component equations,

$$\boldsymbol{d}f = f_{,\alpha} \boldsymbol{\omega}^\alpha \qquad \text{(expansion of } \boldsymbol{d}f \text{ in arbitrary basis),} \tag{9.13a}$$

$$f_{,\alpha} = \partial_\alpha f = \boldsymbol{e}_\alpha[f] \qquad \text{(way to calculate components of } \boldsymbol{d}f\text{),} \tag{9.13b}$$

$$f_{,\alpha} = \partial f / \partial x^\alpha \qquad \text{if } \{\boldsymbol{e}_\alpha\} \text{ is a coordinate basis,}$$

except that they work in arbitrary bases, not just in Lorentz bases. And, as in Lorentz frames, so also in general: the one-form basis $\{\boldsymbol{d}x^\alpha\}$ and the tangent-vector basis $\{\partial/\partial x^\alpha\}$, which are induced into a tangent space by the same coordinate system, are the duals of each other,

$$\langle \boldsymbol{d}x^\alpha, \partial/\partial x^\beta \rangle = \delta^\alpha{}_\beta. \tag{9.14}$$

(See exercise 9.2 for proofs.) Also, most aspects of Cartan's "Exterior Calculus" (parts A, B, C of Box 4.1) are left unaffected by the removal of metric.

§9.5. TENSORS

A tensor $\boldsymbol{S}$, in the absence of Lorentz metric, differs from the tensors of flat, Lorentz spacetime in two ways. (1) $\boldsymbol{S}$ must reside at a specific event $\mathscr{P}_0$, just as any vector or 1-form must. (2) Each slot of $\boldsymbol{S}$ is specific; it will accept either vectors or 1-forms, but not both, because it has no way to convert a 1-form $\tilde{\boldsymbol{u}}$ into a "corresponding

Specificity of tensor slots

vector" $\boldsymbol{u}$ as it sends $\tilde{\boldsymbol{u}}$ through its linear machinery. Thus, if $\boldsymbol{S}$ is a $\binom{1}{2}$ tensor

$$\boldsymbol{S}(\underbrace{\ldots}_{\text{insert 1-form here}},\ \underbrace{\ldots}_{\text{insert vector here}},\ \underbrace{\ldots}_{\text{insert vector here}}), \tag{9.15}$$

then it *cannot* be converted alternatively to a $\binom{2}{1}$ tensor, or a $\binom{3}{0}$ tensor, or a $\binom{0}{3}$ tensor by the procedure of §3.2. In component language, the indices of $\boldsymbol{S}$ cannot be raised and lowered.

Except for these two restrictions (attachment to a specific event; specificity of slots), a tensor $\boldsymbol{S}$ is the same linear machine as ever. And the algebra of component manipulations is the same:

$$S^{\alpha}{}_{\beta\gamma} = \boldsymbol{S}(\boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta}, \boldsymbol{e}_{\gamma}) \qquad (\boldsymbol{S}, \boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta} \text{ must all reside at same event}) \tag{9.16}$$

$$\boldsymbol{S} = S^{\alpha}{}_{\beta\gamma}\boldsymbol{e}_{\alpha} \otimes \boldsymbol{\omega}^{\beta} \otimes \boldsymbol{\omega}^{\gamma}, \tag{9.17}$$

$$\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{u}, \boldsymbol{v}) = S^{\alpha}{}_{\beta\gamma}\sigma_{\alpha}u^{\beta}v^{\gamma}. \tag{9.18}$$

EXERCISES

Exercise 9.1. COMPONENT MANIPULATIONS

Derive equations (9.11c) through (9.11g) from (9.10), (9.11a, b), (9.6), (9.7), and (9.8).

Exercise 9.2. COMPONENTS OF GRADIENT, AND DUALITY OF COORDINATE BASES

In an arbitrary basis, define $f_{,\alpha}$ by the expansion (9.13a). Then combine equations (9.11d) and (9.12) to obtain the method (9.13b) of computing $f_{,\alpha}$. Finally, combine equations (9.12) and (9.13b) to show that the bases $\{\boldsymbol{d}x^{\alpha}\}$ and $\{\partial/\partial x^{\beta}\}$ are the duals of each other.

Exercise 9.3. PRACTICE MANIPULATING TANGENT VECTORS

Let $\mathscr{P}_0$ be the point with coordinates $(x = 0, y = 1, z = 0)$ in a three-dimensional space; and define three curves through $\mathscr{P}_0$ by

$$\mathscr{P}(\lambda) = (\lambda, 1, \lambda),$$

$$\mathscr{P}(\zeta) = (\sin\zeta, \cos\zeta, \zeta),$$

$$\mathscr{P}(\rho) = (\sinh\rho, \cosh\rho, \rho + \rho^3).$$

(a) Compute $(d/d\lambda)f$, $(d/d\zeta)f$, and $(d/d\rho)f$ for the function $f = x^2 - y^2 + z^2$ at the point $\mathscr{P}_0$. (b) Calculate the components of the tangent vectors $d/d\lambda$, $d/d\zeta$, and $d/d\rho$ at $\mathscr{P}_0$, using the basis $\{\partial/\partial x, \partial/\partial y, \partial/\partial z\}$.

Exercise 9.4. MORE PRACTICE WITH TANGENT VECTORS

In a three-dimensional space with coordinates (x, y, z), introduce the vector field $\boldsymbol{v} = y^2\,\partial/\partial x - x\,\partial/\partial z$, and the functions $f = xy$, $g = z^3$. Compute

(*a*) $\boldsymbol{v}[f]$	(*c*) $\boldsymbol{v}[fg]$	(*e*) $\boldsymbol{v}[f^2 + g^2]$
(*b*) $\boldsymbol{v}[g]$	(*d*) $f\boldsymbol{v}[g] - g\boldsymbol{v}[f]$	(*f*) $\boldsymbol{v}\{\boldsymbol{v}[f]\}$

Exercise 9.5. PICTURE OF BASIS 1-FORMS INDUCED BY COORDINATES

In the tangent space of Figure 9.1, draw the basis 1-forms $\boldsymbol{d}\psi$ and $\boldsymbol{d}\chi$ induced by the ψ, χ-coordinate system.

Exercise 9.6. PRACTICE WITH DUAL BASES

In a three-dimensional space with spherical coordinates r, θ, ϕ, one often likes to use, instead of the basis $\partial/\partial r$, $\partial/\partial\theta$, $\partial/\partial\phi$, the basis

$$\boldsymbol{e}_{\hat{r}} = \frac{\partial}{\partial r}, \qquad \boldsymbol{e}_{\hat{\theta}} = \frac{1}{r}\frac{\partial}{\partial\theta}, \qquad \boldsymbol{e}_{\hat{\phi}} = \frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}.$$

(a) What is the 1-form basis $\{\boldsymbol{\omega}^{\hat{r}}, \boldsymbol{\omega}^{\hat{\theta}}, \boldsymbol{\omega}^{\hat{\phi}}\}$ dual to this tangent-vector basis? (b) On the sphere $r = 1$, draw pictures of the bases $\{\partial/\partial r, \partial/\partial\theta, \partial/\partial\phi\}$, $\{\boldsymbol{e}_{\hat{r}}, \boldsymbol{e}_{\hat{\theta}}, \boldsymbol{e}_{\hat{\phi}}\}$, $\{\boldsymbol{d}r, \boldsymbol{d}\theta, \boldsymbol{d}\phi\}$, and $\{\boldsymbol{\omega}^{\hat{r}}, \boldsymbol{\omega}^{\hat{\theta}}, \boldsymbol{\omega}^{\hat{\phi}}\}$.

---

§9.6. COMMUTATORS AND PICTORIAL TECHNIQUES

A vector $\boldsymbol{u}_0$ given only at one point $\mathscr{P}_0$ suffices to compute the derivative $\boldsymbol{u}_0[f] \equiv \partial_{\boldsymbol{u}_0}f$, which is simply a number associated with the point $\mathscr{P}_0$. In contrast, a vector field $\boldsymbol{u}$ provides a vector $\boldsymbol{u}(\mathscr{P})$—which is a differential operator $\partial_{\boldsymbol{u}(\mathscr{P})}$—at each point $\mathscr{P}$ in some region of spacetime. This vector field operates on a function f to produce not just a number, but another function $\boldsymbol{u}[f] \equiv \partial_{\boldsymbol{u}}f$. A second vector field $\boldsymbol{v}$ can perfectly well operate on this new function, to produce yet another function

$$\boldsymbol{v}\{\boldsymbol{u}[f]\} = \partial_{\boldsymbol{v}}(\partial_{\boldsymbol{u}}f).$$

Does this function agree with the result of applying $\boldsymbol{v}$ first and then $\boldsymbol{u}$? Equivalently, does the "commutator"

$$[\boldsymbol{u}, \boldsymbol{v}][f] \equiv \boldsymbol{u}\{\boldsymbol{v}[f]\} - \boldsymbol{v}\{\boldsymbol{u}[f]\} \tag{9.19}$$

Commutator defined

vanish?

The simplest special case is when $\boldsymbol{u}$ and $\boldsymbol{v}$ are basis vectors of a coordinate system, $\boldsymbol{u} = \partial/\partial x^\alpha$, $\boldsymbol{v} = \partial/\partial x^\beta$. Then the commutator does vanish, because partial derivatives always commute:

$$[\partial/\partial x^\alpha, \partial/\partial x^\beta][f] = \partial^2 f/\partial x^\beta\,\partial x^\alpha - \partial^2 f/\partial x^\alpha\,\partial x^\beta = 0.$$

But in general the commutator is nonzero, as one sees from a coordinate-based calculation:

$$\begin{aligned}[\boldsymbol{u}, \boldsymbol{v}]f &= u^\alpha \frac{\partial}{\partial x^\alpha}\left(v^\beta \frac{\partial f}{\partial x^\beta}\right) - v^\alpha \frac{\partial}{\partial x^\alpha}\left(u^\beta \frac{\partial f}{\partial x^\beta}\right) \\ &= \left[(u^\alpha v^\beta{}_{,\alpha} - v^\alpha u^\beta{}_{,\alpha})\frac{\partial}{\partial x^\beta}\right]f.\end{aligned}$$

Commutator of two vector fields is a vector field

Notice however, that the commutator $[\boldsymbol{u}, \boldsymbol{v}]$, like $\boldsymbol{u}$ and $\boldsymbol{v}$ themselves, is a vector field, i.e., a linear differential operator at each event:

$$[\boldsymbol{u}, \boldsymbol{v}] = (\boldsymbol{u}[v^\beta] - \boldsymbol{v}[u^\beta])\frac{\partial}{\partial x^\beta} = (u^\alpha v^\beta{}_{,\alpha} - v^\alpha u^\beta{}_{,\alpha})\frac{\partial}{\partial x^\beta}. \tag{9.20}$$

Such results should be familiar from quantum theory's formalism for angular momentum operators (exercise 9.8).

The three levels of geometry—pictorial, abstract, and component—yield three different insights into the commutator. (1) The abstract expression $[\boldsymbol{u}, \boldsymbol{v}]$ suggests the close connection to quantum theory, and brings to mind the many tools developed there for handling operators. But *recall* that the operators of quantum theory need not be first-order differential operators. The kinetic energy is second order and the potential is zeroth order in the familiar Schrödinger equation. Only first-order operators are vectors. (2) The component expression $u^\alpha v^\beta{}_{,\alpha} - v^\alpha u^\beta{}_{,\alpha}$, valid in any coordinate basis, brings the commutator into the reaches of the powerful tools of index mechanics. (3) The pictorial representation of $[\boldsymbol{u}, \boldsymbol{v}]$ (Box 9.2) reveals its fundamental role as a "closer of curves"—a role that will be important in Chapter 11's analysis of curvature.

Commutator as a "closer of curves"

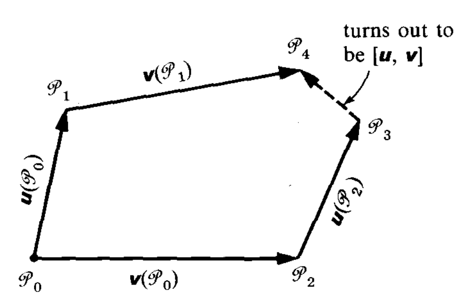
Commutators find application in the distinction between a coordinate-induced basis, $\{\boldsymbol{e}_\alpha\} = \{\partial/\partial x^\alpha\}$, and a noncoordinate basis. Because partial derivatives always commute,

$$[\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta] = [\partial/\partial x^\alpha, \partial/\partial x^\beta] = 0 \text{ in any coordinate basis.} \tag{9.21}$$

Box 9.2 THE COMMUTATOR AS A CLOSER OF QUADRILATERALS

A. Pictorial Representation in Flat Spacetime

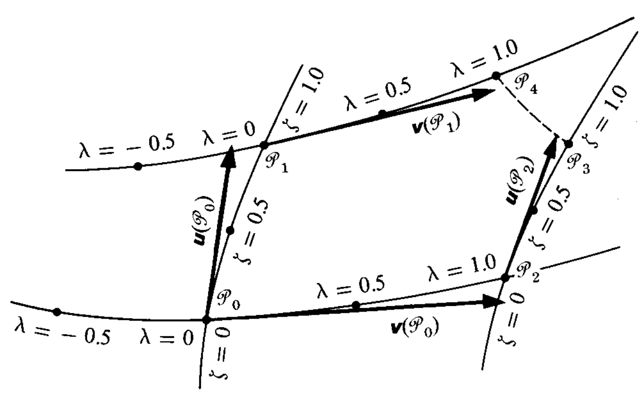
1. For ease of visualization, consider flat spacetime, so the two vector fields $\boldsymbol{u}(\mathscr{P})$ and $\boldsymbol{v}(\mathscr{P})$ can be laid out in spacetime itself.
2. Choose an event $\mathscr{P}_0$ where the commutator $[\boldsymbol{u}, \boldsymbol{v}]$ is to be calculated.
3. Give the names $\mathscr{P}_1$, $\mathscr{P}_2$, $\mathscr{P}_3$, $\mathscr{P}_4$ to the events pictured in the diagram.
4. Then the vector $\mathscr{P}_4 - \mathscr{P}_3$, which measures how much the four-legged curve fails to close, can be expressed in a coordinate basis as



$$\begin{aligned}
\mathscr{P}_4 - \mathscr{P}_3 &= [\boldsymbol{u}(\mathscr{P}_0) + \boldsymbol{v}(\mathscr{P}_1)] - [\boldsymbol{u}(\mathscr{P}_2) + \boldsymbol{v}(\mathscr{P}_0)] \\
&= [\boldsymbol{v}(\mathscr{P}_1) - \boldsymbol{v}(\mathscr{P}_0)] - [\boldsymbol{u}(\mathscr{P}_2) - \boldsymbol{u}(\mathscr{P}_0)] \\
&= (v^{\beta}{}_{,\alpha} u^{\alpha} \boldsymbol{e}_{\beta})_{\mathscr{P}_0} - (u^{\beta}{}_{,\alpha} v^{\alpha} \boldsymbol{e}_{\beta})_{\mathscr{P}_0} + \text{errors} \\
&= [\boldsymbol{u}, \boldsymbol{v}]_{\mathscr{P}_0} + \underbrace{\text{errors}}_{[\text{terms such as } v^{\beta}{}_{,\mu\nu} u^{\mu} u^{\nu} \boldsymbol{e}_{\beta}]}.
\end{aligned}$$

5. Notice that if $\boldsymbol{u}$ and $\boldsymbol{v}$ are halved everywhere, then $[\boldsymbol{u}, \boldsymbol{v}]$ is cut down by a factor of 4, while the error terms in the above go down by a factor of 8. Thus, $[\boldsymbol{u}, \boldsymbol{v}]$ represents accurately the gap in the four-legged curve (*"quadrilateral"*) in the limit where $\boldsymbol{u}$ and $\boldsymbol{v}$ are sufficiently short; i.e., $[\boldsymbol{u}, \boldsymbol{v}]$ *"closes the quadrilateral"* whose edges are the vectors fields $\boldsymbol{u}$ and $\boldsymbol{v}$.

B. Pictorial Representation in Absence of Metric, or in Curved Spacetime with a Metric



1. The same picture must work, but now one dares not (at least initially) lay out the vector fields in spacetime itself. Instead one lays out two families of curves: the curves for which $\boldsymbol{u}(\mathscr{P})$ is the tangent vector; and the curves for which $\boldsymbol{v}(\mathscr{P})$ is the tangent vector.
2. The gap "$\mathscr{P}_4 - \mathscr{P}_3$" in the four-legged curve can be characterized by the difference $f(\mathscr{P}_4) - f(\mathscr{P}_3)$ in the values of an arbitrary function at $\mathscr{P}_4$ and $\mathscr{P}_3$. That difference is, in a coordinate basis,

Box 9.2 ***(continued)***

$$
\begin{aligned}
f(\mathscr{P}_4) - f(\mathscr{P}_3) &= \underbrace{[f(\mathscr{P}_4) - f(\mathscr{P}_1)]}_{\left(f_{,\alpha}v^\alpha + \frac{1}{2}f_{,\alpha\beta}v^\alpha v^\beta\right)_{\mathscr{P}_1}} + \underbrace{[f(\mathscr{P}_1) - f(\mathscr{P}_0)]}_{\left(f_{,\alpha}u^\alpha + \frac{1}{2}f_{,\alpha\beta}u^\alpha u^\beta\right)_{\mathscr{P}_0}} \\
&\quad - \underbrace{[f(\mathscr{P}_2) - f(\mathscr{P}_0)]}_{\left(f_{,\alpha}v^\alpha + \frac{1}{2}f_{,\alpha\beta}v^\alpha v^\beta\right)_{\mathscr{P}_0}} - \underbrace{[f(\mathscr{P}_3) - f(\mathscr{P}_2)]}_{\left(f_{,\alpha}u^\alpha + \frac{1}{2}f_{,\alpha\beta}u^\alpha u^\beta\right)_{\mathscr{P}_2}} \\
&= [(f_{,\alpha}v^\alpha)_{,\beta}u^\beta - (f_{,\alpha}u^\alpha)_{,\beta}v^\beta]_{\mathscr{P}_0} + \text{“cubic errors”} \\
&= [(u^\beta v^\alpha{}_{,\beta} - v^\beta u^\alpha{}_{,\beta})\partial f/\partial x^\alpha]_{\mathscr{P}_0} + \text{“cubic errors”} \\
&= \{[\boldsymbol{u}, \boldsymbol{v}][f]\}_{\mathscr{P}_0} + \text{“cubic errors.”}
\end{aligned}
$$

Here “cubic errors” are cut down by a factor of 8, while $[\boldsymbol{u}, \boldsymbol{v}]f$ is cut down by one of 4, whenever $\boldsymbol{u}$ and $\boldsymbol{v}$ are cut in half.

3. The result

$$f(\mathscr{P}_4) - f(\mathscr{P}_3) = \{[\boldsymbol{u}, \boldsymbol{v}][f]\}_{\mathscr{P}_0} + \text{“cubic errors”}$$

says that $[\boldsymbol{u}, \boldsymbol{v}]$ is a tangent vector at $\mathscr{P}_0$ that describes the separation between the points $\mathscr{P}_3$ and $\mathscr{P}_4$. Its description gets arbitrarily accurate when $\boldsymbol{u}$ and $\boldsymbol{v}$ get arbitrarily short. Thus, $[\boldsymbol{u}, \boldsymbol{v}]$ closes the quadrilateral whose edges are the projections of $\boldsymbol{u}$ and $\boldsymbol{v}$ into spacetime.

C. Philosophy of Pictures

1. Pictures are no substitute for computation. Rather, they are useful for (a) suggesting geometric relationships that were previously unsuspected and that one verifies subsequently by computation; (b) interpreting newly learned geometric results.
2. This usual noncomputational role of pictures permits one to be sloppy in drawing them. No essential new insight was gained in part B over part A, when one carefully moved the tangent vectors into their respective tangent spaces, and permitted only curves to lie in spacetime. Moreover, the original picture (part A) was clearer because of its greater simplicity.
3. This motivates one to draw “sloppy” pictures, with tangent vectors lying in spacetime itself—so long as one keeps those tangent vectors short and occasionally checks the scaling of errors when the lengths of the vectors are halved.

Conversely, if one is given a field of basis vectors ("frame field") $\{\boldsymbol{e}_\alpha(\mathcal{P})\}$, but one does not know whether a coordinate system $\{x^\alpha(\mathcal{P})\}$ exists in which $\{\boldsymbol{e}_\alpha\} = \{\partial/\partial x^\alpha\}$, one can find out by a simple test: calculate all $(4 \times 3)/2 = 6$ commutators $[\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta]$; if they all vanish, then there exists such a coordinate system. If not, there doesn't. Stated more briefly, $\{\boldsymbol{e}_\alpha(\mathcal{P})\}$ *is a coordinate-induced basis if and only if* $[\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta] = 0$ *for all* $\boldsymbol{e}_\alpha$ *and* $\boldsymbol{e}_\beta$. (See exercise 9.9 for proof; see §11.5 for an important application.) Coordinate-induced bases are sometimes called *"holonomic."* In an *"anholonomic* basis" (noncoordinate basis), one defines the commutation coefficients $c_{\mu\nu}{}^\alpha$ by

Vanishing commutator: a test for coordinate bases

Commutation coefficients defined

$$[\boldsymbol{e}_\mu, \boldsymbol{e}_\nu] = c_{\mu\nu}{}^\alpha \boldsymbol{e}_\alpha. \tag{9.22}$$

They enter into the component formula for the commutator of arbitrary vector fields $\boldsymbol{u}$ and $\boldsymbol{v}$:

$$[\boldsymbol{u}, \boldsymbol{v}] = (\boldsymbol{u}[v^\beta] - \boldsymbol{v}[u^\beta] + u^\mu v^\nu c_{\mu\nu}{}^\beta)\boldsymbol{e}_\beta \tag{9.23}$$

(see exercise 9.10).

Physicists' notation vs. mathematicians' notation

[*Warning!* In notation for functions and fields, mathematicians and physicists often use the same symbols to mean contradictory things. The physicist may write ℓ when considering the length of some critical component in an instrument he is designing, then switch to $\ell(T)$ when he begins to analyze its response to temperature changes. Thus ℓ is a number, whereas $\ell(T)$ is a function. The mathematician, in contrast, will write f for a function that he may be considering as an element in some infinite-dimensional function space. Once the function is supplied with an argument, he then contemplates $f(x)$, which is merely a number: the value of f at the point x. Caught between these antithetical rituals of the physics and mathematics sects, the authors have adopted a clear policy: vacillation. Usually physics-sect statements, like "On a curve $\mathcal{P}(\lambda)\ldots$," are used; and the reader can translate them himself into mathematically precise language: "Consider a curve $\mathcal{C}$ on which a typical point is $\mathcal{P} = \mathcal{C}(\lambda)$; on this curve ...." But on occasion the reader will encounter a pedantic-sounding paragraph written in mathematics-sect jargon (Example: Box 23.3). Such paragraphs deal with concepts and relationships so complex that standard physics usage would lead to extreme confusion. They also should prevent the reader from becoming so conditioned to physics usage that he is allergic to the mathematical literature, where great advantages of clarity and economy of thought are achieved by consistent reliance on wholly unambiguous notation.]

EXERCISES

Exercise 9.7. PRACTICE WITH COMMUTATORS

Compute the commutator $[\boldsymbol{e}_{\hat\theta}, \boldsymbol{e}_{\hat\phi}]$ of the vector fields

$$\boldsymbol{e}_{\hat\theta} = \frac{1}{r}\frac{\partial}{\partial\theta}, \qquad \boldsymbol{e}_{\hat\phi} = \frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}.$$

Express your result as a linear combination of $\boldsymbol{e}_{\hat\theta}$ and $\boldsymbol{e}_{\hat\phi}$.

Exercise 9.8. ANGULAR MOMENTUM OPERATORS

In Cartesian coordinates of three-dimensional Euclidean space, one defines three *"angular-momentum operators"* (vector fields) $\boldsymbol{L}_j$ by

$$\boldsymbol{L}_j \equiv \epsilon_{jkl} x^k (\partial/\partial x^l).$$

Draw a picture of these three vector fields. Calculate their commutators both pictorially and analytically.

Exercise 9.9. COMMUTATORS AND COORDINATE-INDUCED BASES

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vector fields in spacetime. Show that in some neighborhood of any given point there exists a coordinate system for which

$$\boldsymbol{u} = \partial/\partial x^1, \qquad \boldsymbol{v} = \partial/\partial x^2,$$

if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent and commute:

$$[\boldsymbol{u}, \boldsymbol{v}] = 0.$$

First make this result plausible from the second figure in Box 9.2; then prove it mathematically. *Note:* this result can be generalized to four arbitrary vector fields $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$. There exists a coordinate system in which $\boldsymbol{e}_\alpha = \partial/\partial x^\alpha$ if and only if $\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ are linearly independent and $[\boldsymbol{e}_\mu, \boldsymbol{e}_\nu] = 0$ for all pairs $\boldsymbol{e}_\mu$, $\boldsymbol{e}_\nu$.

Exercise 9.10. COMPONENTS OF COMMUTATOR IN NON-COORDINATE BASIS

Derive equation (9.23).

Exercise 9.11. LIE DERIVATIVE

The "Lie derivative" of a vector field $\boldsymbol{v}(\mathscr{P})$ along a vector field $\boldsymbol{u}(\mathscr{P})$ is defined by

$$\pounds_{\boldsymbol{u}} \boldsymbol{v} \equiv [\boldsymbol{u}, \boldsymbol{v}]. \tag{9.24}$$

Draw a space-filling family of curves (a *"congruence"*) on a sheet of paper. Draw an arbitrary vector $\boldsymbol{v}$ at an arbitrary point $\mathscr{P}_0$ on the sheet. Transport that vector along the curve through $\mathscr{P}_0$ by means of the *"Lie transport law"* $\pounds_{\boldsymbol{u}} \boldsymbol{v} = 0$, where $\boldsymbol{u} = d/dt$ is the tangent to the curve. Draw the resulting vector $\boldsymbol{v}$ at various points $\mathscr{P}(t)$ along the curve.

Exercise 9.12. A CHIP OFF THE OLD BLOCK

(a) Prove the Jacobi identity

$$[\boldsymbol{u}, [\boldsymbol{v}, \boldsymbol{w}]] + [\boldsymbol{v}, [\boldsymbol{w}, \boldsymbol{u}]] + [\boldsymbol{w}, [\boldsymbol{u}, \boldsymbol{v}]] = 0 \tag{9.25}$$

by picking out all terms of the form $\partial_{\boldsymbol{u}}\, \partial_{\boldsymbol{v}}\, \partial_{\boldsymbol{w}}$, showing that they add to zero, and arguing from symmetry that all other terms, e.g., $\partial_{\boldsymbol{w}}\, \partial_{\boldsymbol{u}}\, \partial_{\boldsymbol{v}}$ terms, must similarly cancel.

(b) State this identity in index form.

(c) Draw a picture corresponding to this identity (see Box 9.2).

§9.7. MANIFOLDS AND DIFFERENTIAL TOPOLOGY

Spacetime is not the only arena in which the ideas of this chapter can be applied. Points, curves, vectors, 1-forms, and tensors exist in any "differentiable manifold."

Their use to study differentiable manifolds constitutes a branch of mathematics called *"differential topology"*—hence the title of this chapter.

The mathematician usually begins his development of differential topology by introducing some very primitive concepts, such as sets and topologies of sets, by building a fairly elaborate framework out of them, and by then using that framework to define the concept of a differentiable manifold. But most physicists are satisfied with a more fuzzy, intuitive definition of manifold: roughly speaking, an n-dimensional *differentiable manifold* is a set of "points" tied together continuously and differentiably, so that the points in any sufficiently small region can be put into a one-to-one correspondence with an open set of points of R^n. [R^n is the number space of n dimensions, i.e., the space of ordered n-tuples $(x^1, x^2, \ldots, x^n)$.] That correspondence furnishes a coordinate system for the neighborhood.

Differentiable manifold "defined"

A few examples will convey the concept better than this definition. Elementary examples (Euclidean 3-spaces, the surface of a sphere) bring to mind too many geometric ideas from richer levels of geometry; so one is forced to contemplate something more complicated. Let R^3 be a three-dimensional number space with the usual advanced-calculus ideas of continuity and differentiability. Points $\boldsymbol{\xi}$ of R^3 are triples, $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$, of real numbers. Let a *ray* $\mathscr{P}$ in R^3 be any half-line from the origin consisting of all $\boldsymbol{\xi}$ of the form $\boldsymbol{\xi} = \lambda\boldsymbol{\eta}$ for some fixed $\boldsymbol{\eta} \neq 0$ and for all positive real numbers $\lambda > 0$. (See Figure 9.3.) A good example of a differentiable manifold then is the set S^2 of all distinct rays. If f is a real-valued function with a specific value $f(\mathscr{P})$ for any ray $\mathscr{P}$ [so one writes $f\colon S^2 \longrightarrow R\colon \mathscr{P} \longrightarrow f(\mathscr{P})$], it should be intuitively (or even demonstrably) clear that we can define what we mean by saying f is continuous or differentiable. In this sense S^2 itself is continuous and differentiable. Thus S^2 is a manifold, and the rays $\mathscr{P}$ are the points of S^2. There are many other manifolds that differential topology finds indistinguishable from S^2. The simplest is the two-dimensional spherical surface (2-sphere), which is the standard representation of S^2; it is the set of points $\boldsymbol{\xi}$ of R^3 satisfying $(\xi_1)^2 + (\xi_2)^2 + (\xi_3)^2 = 1$. Clearly a different point $\mathscr{P}$ of S^2 (one ray in R^3) intersects each point of this standard 2-sphere surface, and the correspondence is continuous and differentiable in either direction (ray to point; point to ray). The same is true for any ellipsoidal surface in R^3 enclosing the origin, and for any other surface enclosing the origin that has

Examples of differentiable manifolds

The manifold S^2

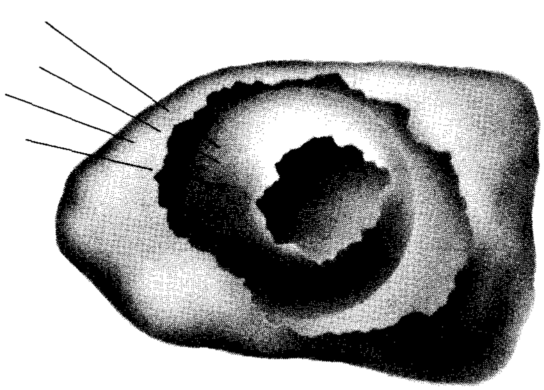


Figure 9.3.
Three different representations of the differentiable manifold S^2. The first is the set of all rays emanating from the origin; the second is the sphere they intersect; the third is an odd-shaped, closed surface that each ray intersects precisely once.

a different ray through each point of itself. They each embody the same global continuity and differentiability concepts, and represent the same abstract differentiable manifold S^2, the 2-sphere. They, and the bundle of rays we started with, all have the same geometric properties at this rudimentary level of geometry. A two-dimensional manifold that has a different geometric structure at this level (a different "differentiable structure") is the torus T^2, the surface of a donut. There is no way to imbed this surface smoothly in R^3 so that a distinct ray $\mathcal{P} \in S^2$ intersects each of its points; there is no invertible and differentiable correspondence between T^2 and S^2.

The manifold T^2

The manifold $SO(3)$ (rotation group)

Another example of a manifold is the rotation group $SO(3)$, whose points $\mathcal{P}$ are all the 3×3 orthogonal matrices of unit determinant, so $\mathcal{P} = \|P_{ij}\|$ with $\mathcal{P}^T\mathcal{P} = 1$ and $\det \mathcal{P} = 1$. This is a three-dimensional space (one often uses the three Euler-angle parameters in computations), where differential ideas (e.g., angular velocity) are employed; hence, it is a manifold. So is the Lorentz group.

The differentiability of a manifold (i.e., the possibility of defining differentiable functions on it) permits one to introduce coordinate systems locally, if not globally, and also curves, tangent spaces, tangent vectors, 1-forms, and tensors, just as is done for spacetime. *But* the mere fact that a manifold is differentiable does not mean that such concepts as geodesics, parallel transport, curvature, metric, or length exist in it. These are additional layers of structure possessed by some manifolds, but not by all. Roughly speaking, every manifold has smoothness properties and topology, but without additional structure it is shapeless and sizeless.

Affine geometry and Riemannian geometry defined

That branch of mathematics which adds geodesics, parallel transport, and curvature (shape) to a manifold is called *affine geometry;* that branch which adds a metric is called *Riemannian geometry.* They will be studied in the next few chapters.

EXERCISES

EXERCISES ON THE ROTATION GROUP

As the exposition of differential geometry becomes more and more sophisticated in the following chapters, the exercises will return time and again to the rotation group as an example of a manifold. Then, in Box 30.1, the results developed in these exercises will be used to analyze the "Mixmaster universe," which is a particularly important cosmological solution to Einstein's field equation.

Before working these exercises, the reader may wish to review the Euler-angle parametrization for rotation matrices, as treated, e.g., on pp. 107–109 of Goldstein (1959).

Exercise 9.13. ROTATION GROUP: GENERATORS

Let $\mathcal{K}_l$ be three 3×3 matrices whose components are $(K_l)_{mn} = \epsilon_{lmn}$.

(a) Display the matrices $\mathcal{K}_1$, $(\mathcal{K}_1)^2$, $(\mathcal{K}_1)^3$, and $(\mathcal{K}_1)^4$.

(b) Sum the series

$$\mathcal{R}_x(\theta) \equiv \exp(\mathcal{K}_1\theta) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} (\mathcal{K}_1)^n. \tag{9.26}$$

Show that $\mathcal{R}_x(\theta)$ is a rotation matrix and that it produces a rotation through an angle θ about the x-axis.

(c) Show similarly that $\mathcal{R}_z(\phi) = \exp(\mathcal{K}_3\phi)$ and $\mathcal{R}_y(\chi) = \exp(\mathcal{K}_2\chi)$ are rotation matrices, and that they produce rotations through angles ϕ and χ about the z- and y-axes, respectively.

(d) Explain why $\mathcal{P} = \mathcal{R}_z(\psi)\mathcal{R}_x(\theta)\mathcal{R}_z(\phi)$ defines the Euler-angle coordinates, ψ, θ, ϕ for the generic element $\mathcal{P} \in SO(3)$ of the rotation group.

(e) Let $\mathcal{C}$ be the curve $\mathcal{P} = \mathcal{R}_z(t)$ through the identity matrix, $\mathcal{C}(0) = \mathcal{I} \in SO(3)$. Show that its tangent, $(d\mathcal{C}/dt)(0) \equiv \dot{\mathcal{C}}(0)$ does not vanish by computing $\dot{\mathcal{C}}(0)f_{12}$, where f_{12} is the function $f_{12}(\mathcal{P}) = P_{12}$, whose value is the 12 matrix element of $\mathcal{P}$.

(f) Define a vector field $\boldsymbol{e}_3$ on $SO(3)$ by letting $\boldsymbol{e}_3(\mathcal{P})$ be the tangent (at $t = 0$) to the curve $\mathcal{C}(t) = \mathcal{R}_z(t)\mathcal{P}$ through $\mathcal{P}$. Show that $\boldsymbol{e}_3(\mathcal{P})$ is nowhere zero. *Note:* $\boldsymbol{e}_3(\mathcal{P})$ is called the "generator of rotations about the z-axis," because it points from $\mathcal{P}$ toward neighboring rotations, $\mathcal{R}_z(t)\mathcal{P}$, which differ from $\mathcal{P}$ by a rotation about the z-axis.

(g) Show that $\boldsymbol{e}_3 = (\partial/\partial\psi)_{\theta\phi}$.

(h) Derive the following formulas, valid for $t \ll 1$:

$$\mathcal{R}_x(t)\mathcal{R}_z(\psi)\mathcal{R}_x(\theta)\mathcal{R}_z(\phi) = \mathcal{R}_z(\psi - t\sin\psi\cot\theta)\mathcal{R}_x(\theta + t\cos\psi)\mathcal{R}_z(\phi + t\sin\psi/\sin\theta);$$
$$\mathcal{R}_y(t)\mathcal{R}_z(\psi)\mathcal{R}_x(\theta)\mathcal{R}_z(\phi) = \mathcal{R}_z(\psi + t\cos\psi\cot\theta)\mathcal{R}_x(\theta + t\sin\psi)\mathcal{R}_z(\phi - t\cos\psi/\sin\theta).$$

(i) Define $\boldsymbol{e}_1(\mathcal{P})$ and $\boldsymbol{e}_2(\mathcal{P})$ to be the tangent vectors (at $t = 0$) to the curves $\mathcal{C}(t) = \mathcal{R}_x(t)\mathcal{P}$ and $\mathcal{C}(t) = \mathcal{R}_y(t)\mathcal{P}$, respectively. Show that

$$\boldsymbol{e}_1 = \cos\psi\frac{\partial}{\partial\theta} - \sin\psi\left(\cot\theta\frac{\partial}{\partial\psi} - \frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right),$$

$$\boldsymbol{e}_2 = \sin\psi\frac{\partial}{\partial\theta} + \cos\psi\left(\cot\theta\frac{\partial}{\partial\psi} - \frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right).$$

$\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are the "generators of rotations about the x- and y-axes."

Exercise 9.14. ROTATION GROUP: STRUCTURE CONSTANTS

Use the three vector fields constructed in the last exercise,

$$\boldsymbol{e}_1 = \cos\psi\frac{\partial}{\partial\theta} - \sin\psi\left(\cot\theta\frac{\partial}{\partial\psi} - \frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right),$$
$$\boldsymbol{e}_2 = \sin\psi\frac{\partial}{\partial\theta} + \cos\psi\left(\cot\theta\frac{\partial}{\partial\psi} - \frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right), \tag{9.27}$$
$$\boldsymbol{e}_3 = \frac{\partial}{\partial\psi},$$

as basis vectors for the manifold of the rotation group. The above equations express this "basis of generators" in terms of the Euler-angle basis. Show that the commutation coefficients for this basis are

$$c_{\alpha\beta}{}^{\gamma} = -\epsilon_{\alpha\beta\gamma}, \tag{9.28}$$

independently of location $\mathcal{P}$ in the rotation group. These coefficients are also called the *structure constants* of the rotation group.

2

CHAPTER **10**

AFFINE GEOMETRY: GEODESICS, PARALLEL TRANSPORT, AND COVARIANT DERIVATIVE

Galilei's Principle of Inertia is sufficient in itself to prove conclusively that the world is affine in character.

HERMANN WEYL

§10.1. GEODESICS AND THE EQUIVALENCE PRINCIPLE

This chapter is entirely Track 2.
Chapter 9 is necessary preparation for it.
It will be needed as preparation for
(1) Chapters 11–13 (differential geometry; Newtonian gravity),
(2) the second half of Chapter 14 (calculation of curvature), and
(3) the details, but not the message, of Chapter 15 (Bianchi identities).

Free fall is the "natural state of motion," so natural, in fact, that *the path through spacetime of a freely falling, neutral test body is independent of its structure and composition* (the "weak equivalence principle" of Einstein, Eötvös, Dicke; see Box 1.2 and §38.3).

Picture spacetime as filled with free-fall trajectories. Pick an event. Pick a velocity there. They determine a unique trajectory.

Freely falling particles and their clocks

Be more precise. Ask for the maximum amount of information tied up in each trajectory. Is it merely the sequence of points along which the test body falls? No; there is more. Each test body can carry a clock with itself (same kind of clock—"good" clock in sense of Figure 1.9—regardless of structure or composition of test body). The clock ticks as the body moves, labeling each event on its trajectory with a number: the time λ the body was there. Result: the free-fall trajectory is not just a sequence of points; it is a parametrized sequence, a *"curve"* $\mathscr{P}(\lambda)$.

But is the parametrization unique? Not entirely. Quite arbitrary are (1) the choice of time origin, $\mathscr{P}(0)$; and (2) the units (centimeters, seconds, furlongs, . . .) in which clock time λ is measured. Hence, λ is unique only up to linear transformations

$$\lambda_{\text{new}} = a\lambda_{\text{old}} + b; \tag{10.1}$$

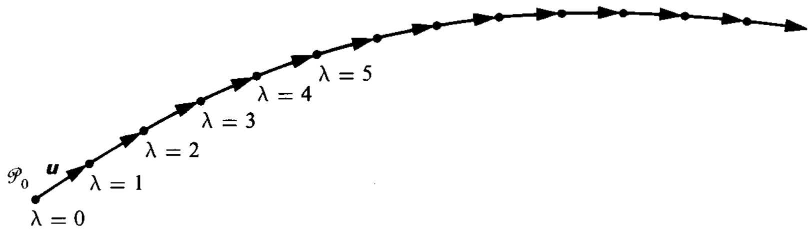


Figure 10.1.
A geodesic viewed as a rule for "straight-on parallel transport." Pick an event $\mathscr{P}_0$ and a tangent vector $\boldsymbol{u} = d/d\lambda$ there. Construct the unique geodesic $\mathscr{P}(\lambda)$ that (1) passes through $\mathscr{P}_0$: $\mathscr{P}(0) = \mathscr{P}_0$; and (2) has $\boldsymbol{u}$ as its tangent vector there: $(d\mathscr{P}/d\lambda)_{\lambda=0} = \boldsymbol{u}$. This geodesic can be viewed as a rule for picking up $\boldsymbol{u}$ from $\mathscr{P}(0)$ and laying it down again at its tip, $\mathscr{P}(1)$, in as straight a manner as possible,

$$\boldsymbol{u}_{\lambda=1} = (d\mathscr{P}/d\lambda)_{\lambda=1};$$

and for then picking it up and laying it down as straight as possible again at $\mathscr{P}(2)$,

$$\boldsymbol{u}_{\lambda=2} = (d\mathscr{P}/d\lambda)_{\lambda=2};$$

etc. This sequence of "straight as possible," "tail-on-tip" transports gives meaning to the idea that $(d\mathscr{P}/d\lambda)_{\lambda=17}$ and $\boldsymbol{u} = (d\mathscr{P}/d\lambda)_{\lambda=0}$ are "the same vector" at different points along the geodesic; or, equivalently, that one has been obtained from the other by "*straight-on parallel transport.*"

b ("new origin of clock time") is a number independent of location on this specific free-fall trajectory, and a ("ratio of new units to old") is also.

In the curved spacetime of Einstein (and in that of Cartan-Newton, Chapter 12), these parametrized free-fall trajectories are the straightest of all possible curves. Consequently, one gives these trajectories the same name, "*geodesics,*" that mathematicians use for the straight lines of a curved manifold; and like the mathematicians, one uses the name "*affine parameter*" for the parameter λ along a free-fall geodesic. Equation (10.1) then says "the affine parameter of a geodesic is unique up to linear transformations."

Geodesic defined as a free-fall trajectory

The affine parameter ("clock time") along a geodesic has nothing to do, à priori, with any metric. It exists even in the absence of metric (e.g., in Cartan-Newtonian spacetime). It gives one a method for comparing the separation between events on a geodesic ($\mathscr{B}$ and $\mathscr{A}$ are "twice as far apart" as $\mathscr{R}$ and $\mathscr{Q}$ if $[\lambda_{\mathscr{B}} - \lambda_{\mathscr{A}}] = 2[\lambda_{\mathscr{R}} - \lambda_{\mathscr{Q}}]$). But the affine parameter measures relative separations only along its own geodesic. It has no means of reaching off the geodesic.

Affine parameter defined as clock time along free-fall trajectory

The above features of geodesics, and others, are summarized in Figure 10.1 and Box 10.1.

§10.2. PARALLEL TRANSPORT AND COVARIANT DERIVATIVE: PICTORIAL APPROACH

Two test bodies, initially falling through spacetime on parallel, neighboring geodesics, get pushed toward each other or apart by tidal gravitational forces (spacetime curvature). To quantify this statement, one must quantify the concepts of "parallel" and "rate of acceleration away from each other." Begin with parallelism.

Box 10.1 GEODESICS

Geodesic in brief

Give point, give tangent vector; get unique, affine-parametrized curve ("geodesic").

Geodesic: in context of gravitation physics

World line of a neutral test particle ("Einstein's geometric theory of gravity"; also "Cartan's translation into geometric terms of Newton's theory of gravity"):

(1) "given point": some event on this world line;

(2) "given vector": vector ("displacement per unit increase of parameter") tangent to world line at instant defined by that event;

(3) "unique curve": every neutral test particle with a specified initial position and a specified initial velocity follows the same world line, regardless of its composition and regardless of its mass (small; test mass!; "weak equivalence principle of Einstein-Eötvös-Dicke");

(4) "affine parameter": in Cartan-Newton theory, Newton's "universal time" (which is measured by "good" clocks); in the real physical world, "proper time" (as measured by a "good" clock) along a timelike geodesic;

(5) "parametrized curve": (a) affine parameter unique up to a transformation of the form $\lambda \longrightarrow a\lambda + b$, where a and b are constants (no arbitrariness along a given geodesic other than zero of parameter and unit of parameter); or equivalently (b) given any three events $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ on the geodesic, one can find by well-determined physical construction ("clocking") a unique fourth event $\mathcal{D}$ on the geodesic such that $(\lambda_{\mathcal{D}} - \lambda_{\mathcal{C}})$ is equal to $(\lambda_{\mathcal{B}} - \lambda_{\mathcal{A}})$; or equivalently (c) [differential version] given a tangent vector with components $(dx^\alpha/d\lambda)_{\mathcal{A}}$ at point $\mathcal{A}$, one can find by physical construction (again "clocking") "the same tangent vector" at point $\mathcal{C}$ with uniquely determined components $(dx^\alpha/d\lambda)_{\mathcal{C}}$ (vector "equal"; components ordinarily not equal because of twisting and turning of arbitrary base vectors between $\mathcal{A}$ and $\mathcal{C}$).

Comparison of vectors at different events by parallel transport

Consider two neighboring events $\mathcal{A}$ and $\mathcal{B}$ connected by a curve $\mathcal{P}(\lambda)$. A vector $\boldsymbol{v}_{\mathcal{A}}$ lies in the tangent space at $\mathcal{A}$, and a vector $\boldsymbol{v}_{\mathcal{B}}$ lies in the tangent space at $\mathcal{B}$. How can one say whether $\boldsymbol{v}_{\mathcal{A}}$ and $\boldsymbol{v}_{\mathcal{B}}$ are parallel, and how can one compare their lengths? The equivalence principle gives an answer: an observer travels (using rocket power as necessary) through spacetime along the world line $\mathcal{P}(\lambda)$. He carries the vector $\boldsymbol{v}_{\mathcal{A}}$ with himself as he moves, and he uses flat-space Newtonian or Minkowskian standards to keep it always unchanging (flat-space physics is valid locally according to the equivalence principle!). On reaching event $\mathscr{B}$ the observer compares his "parallel-transported vector" $\boldsymbol{v}_{\mathscr{A}}$ with the vector $\boldsymbol{v}_{\mathscr{B}}$. If they are identical, then the original vector $\boldsymbol{v}_{\mathscr{A}}$ was (by definition) parallel to $\boldsymbol{v}_{\mathscr{B}}$, and they had the same length. (No metric means no way to quantify length; nevertheless, parallel transport gives a way to compare length!)

Parallel transport defined using inertial guidance systems and equivalence principle

The equivalence principle entered this discussion in a perhaps unfamiliar way, applied to an observer who may be accelerated, rather than to one who is freely falling. But one cannot evade a basic principle by merely confronting it with an intricate application. (Ingenious perpetual-motion machines are as impossible as simpleminded ones!) The equivalence principle states that no local measurement that is insensitive to gravitational tidal forces can detect any difference whatsoever between flat and curved spacetime. The spaceship navigator has an inertial guidance system (accelerometers, gyroscopes, computers) capable of preserving an inertial reference frame in flat spacetime; and in flat spacetime it can compute the attitude and velocity of any object in the spaceship relative to a given inertial frame. The purchaser may specify whether he wants a guidance computer programmed with the laws of zero-gravity Newtonian mechanics, or with those of special-relativity physics. Use this same guidance system—including the same computer program—in curved spacetime. A vector is being parallel transported if the guidance system's computer says it is not changing.

Result of parallel transport depends on route

Will the result of transport in this way be independent of the curve used to link $\mathscr{A}$ and $\mathscr{B}$? Clearly yes, in gravity-free spacetime, since this is a principal performance criterion that the purchaser of an inertial guidance system can demand of the manufacturer. But in a curved spacetime, the answer is "NO!" If $\boldsymbol{v}_{\mathscr{A}}$ agrees with $\boldsymbol{v}_{\mathscr{B}}$ after parallel transport along one curve, it need not agree with $\boldsymbol{v}_{\mathscr{B}}$ after parallel transport along another. Spacetime curvature produces discrepancies. But one is not ready to study and quantify those discrepancies (Chapter 11), until one has developed the mathematical formalism of parallel transport, which, in turn, cannot be done until one has made precise the "flat-space standards for keeping the vector $\boldsymbol{v}_{\mathscr{A}}$ always unchanging" as it is transported along a curve.

Schild's ladder for performing parallel transport; its consequences

The flat-space standards are made precise in Box 10.2. They lead to (1) a "Schild's ladder" construction for performing parallel transport; (2) the concept "covariant derivative," $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}$, of a vector field $\boldsymbol{v}$ along a curve with tangent $\boldsymbol{u}$; (3) the "equation of motion" $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0$ for a geodesic, which states that "a geodesic parallel transports its own tangent vector along itself;" and (4) a link between the tangent spaces at adjacent events (Figure 10.2).

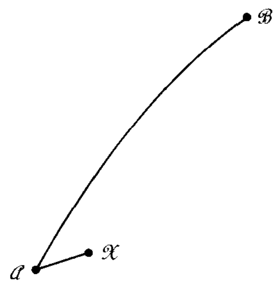
§10.3. PARALLEL TRANSPORT AND COVARIANT DERIVATIVE: ABSTRACT APPROACH

From the "Schild's ladder" construction of Box 10.2, one learns the following properties of spacetime's covariant derivative:

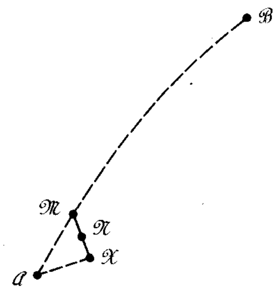
(*continued on page 252*)

Box 10.2 FROM GEODESICS TO PARALLEL TRANSPORT TO COVARIANT DIFFERENTIATION TO GEODESICS TO . . .

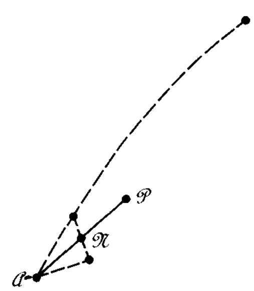
"Parallel transport" as defined by geodesics



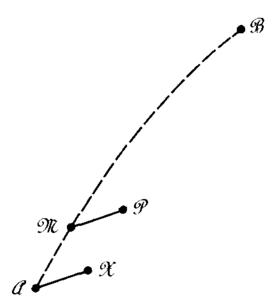
A. Transport any sufficiently short stretch of a curve $\mathcal{A}\mathcal{X}$ (i.e., any tangent vector) parallel to itself along curve $\mathcal{A}\mathcal{B}$ to point $\mathcal{B}$ as follows:



1. Take some point $\mathcal{M}$ along $\mathcal{A}\mathcal{B}$ close to $\mathcal{A}$. Take geodesic $\mathcal{X}\mathcal{M}$ through $\mathcal{X}$ and $\mathcal{M}$. Take any affine parametrization λ of $\mathcal{X}\mathcal{M}$ and define a unique point $\mathcal{N}$ by the condition $\lambda_{\mathcal{N}} = \frac{1}{2}(\lambda_{\mathcal{X}} + \lambda_{\mathcal{M}})$ ("equal stretches of time in $\mathcal{X}\mathcal{N}$ and $\mathcal{N}\mathcal{M}$").

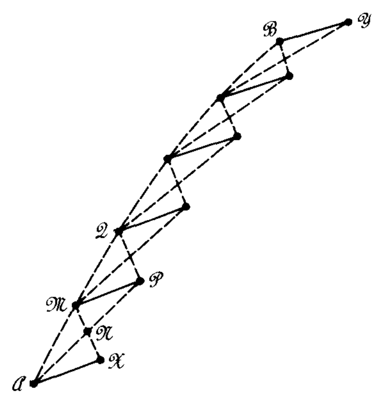


2. Take geodesic that starts at $\mathcal{A}$ and passes through $\mathcal{N}$, and extend it by an equal parameter increment to point $\mathcal{P}$.



3. Curve $\mathcal{M}\mathcal{P}$ gives vector $\mathcal{A}\mathcal{X}$ as propagated parallel to itself from $\mathcal{A}$ to $\mathcal{M}$ (for sufficiently short $\mathcal{A}\mathcal{X}$ and $\mathcal{A}\mathcal{M}$). This construction certainly yields parallel transport in flat spacetime (Newtonian or Einsteinian). Moreover, it is local (vectors $\mathcal{A}\mathcal{X}$, $\mathcal{A}\mathcal{M}$, etc., very short). Therefore, it must work even in curved spacetime. (It embodies the equivalence principle.)

4. Repeat process over and over, and eventually end up with $\mathcal{A}\mathcal{X}$ propagated parallel to itself from $\mathcal{A}$ to $\mathcal{B}$. Call this construction "Schild's Ladder," from Schild's (1970) similar construction. [See also Ehlers, Pirani, and Schild (1972).] Note that curve $\mathcal{A}\mathcal{B}$ need not be a geodesic. There is no requirement that $\mathcal{M}\mathcal{Q}$ be the straight-on continuation of $\mathcal{A}\mathcal{M}$ similar to the geodesic requirement in the "cross-brace" that $\mathcal{N}\mathcal{P}$ be the straight-on continuation of $\mathcal{A}\mathcal{N}$.



"Schild's Ladder"

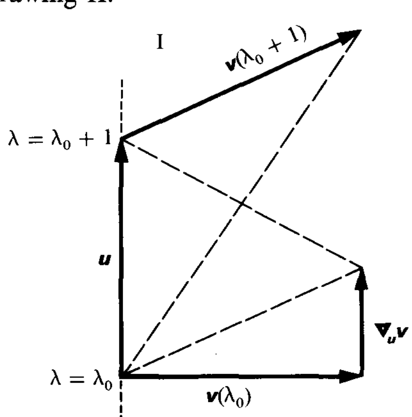
5. Result of propagating $\mathcal{A}\mathcal{X}$ parallel to itself from $\mathcal{A}$ to $\mathcal{B}$ depends on choice of world line $\mathcal{A}\mathcal{B}$ ("evidence of curvature of spacetime").

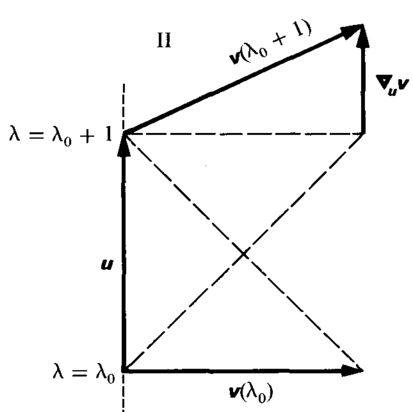
B. Ask how rapidly a vector field $\boldsymbol{v}$ is changing along a curve with tangent vector $\boldsymbol{u} = d/d\lambda$. The answer, $d\boldsymbol{v}/d\lambda \equiv \boldsymbol{\nabla_u v} \equiv$ "rate of change of $\boldsymbol{v}$ with respect to λ" $\equiv$ "covariant derivative of $\boldsymbol{v}$ along $\boldsymbol{u}$," is constructed by the following obvious procedure: (1) Take $\boldsymbol{v}$ at $\lambda = \lambda_0 + \varepsilon$. (2) Parallel transport it back to $\lambda = \lambda_0$. (3) Calculate how much it differs from $\boldsymbol{v}$ there. (4) Divide by ε (and take limit as $\varepsilon \longrightarrow 0$):

"Covariant differentiation" as defined by parallel transport

$$\boldsymbol{\nabla_u v} = \lim_{\varepsilon \to 0} \left\{ \frac{[\boldsymbol{v}(\lambda_0 + \varepsilon)]_{\text{parallel transported to } \lambda_0} - \boldsymbol{v}(\lambda_0)}{\varepsilon} \right\}.$$

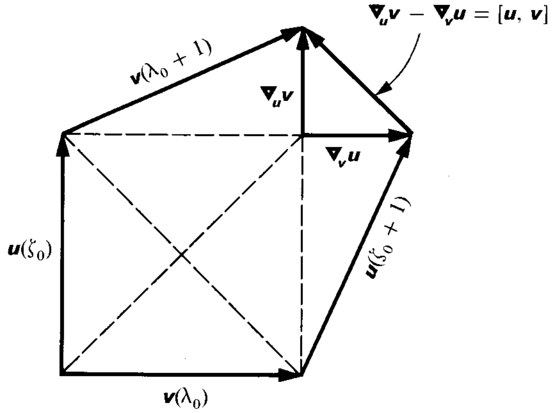
If $\boldsymbol{u} = d/d\lambda$ is short compared to scale of inhomogeneities in the vector field $\boldsymbol{v}$, then $\boldsymbol{\nabla_u v}$ can be read directly off drawing I, or, equally well, off drawing II.





Box 10.2 (*continued*)

"Symmetry" of covariant differentiation



C. Take two vector fields. Combine into one the two diagrams for $\nabla_u v$ and $\nabla_v u$. Thereby discover that $\nabla_u v - \nabla_v u$ is the vector by which the ***v-u-v-u*** quadrilateral fails to close—i.e. (see Box 9.2), it is the commutator $[u, v]$: $\nabla_u v - \nabla_v u = [u, v]$.

Terminology: ∇ is said to be a *"symmetric"* or *"torsion-free"* covariant derivative when $\nabla_u v - \nabla_v u = [u, v]$. Other types of covariant derivatives, as studied by mathematicians, have no relevance for any gravitation theory based on the equivalence principle.

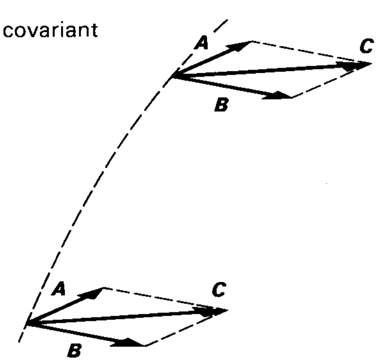
Chain rule for covariant differentiation

D. The "take-the-difference" and "take-the-limit" process used to define $\nabla_u v$ guarantees that it obeys the usual rule for differentiating products:

$$\nabla_u(\underset{\text{scalar field}}{f}\,\underset{\text{vector field}}{v}) = f\nabla_u v + \underbrace{(u[f])}_{\substack{\text{“derivative of } f \text{ along } u\text{,” denoted } \partial_u f \text{ in first}\\ \text{part of book; actually equal to } df/d\lambda \text{ if } u = d/d\lambda\text{;}\\ \text{also sometimes denoted } \nabla_u f.}} v$$

(for proof, see exercise 10.2.)

Additivity for covariant differentiation



E. In the real physical world, be it Newtonian or relativistic, parallel transport of a triangle cannot break its legs apart: (1) ***A***, ***B***, ***C*** initially such that $A + B = C$; (2) ***A***, ***B***, ***C*** each parallel transported with himself by freely falling (inertial) observer; (3) then $A + B = C$ always. Any other result would violate the equivalence principle!

1. Consequence of this (as seen by following through definition of covariant derivative, and by noting that any vector $\boldsymbol{u}$ can be regarded as the tangent vector to a freely falling world line):

$$\boldsymbol{\nabla}_{\boldsymbol{u}}(\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} + \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}$$

for any vector $\boldsymbol{u}$ and vector fields $\boldsymbol{v}$ and $\boldsymbol{w}$.

2. Consequence of this, combined with symmetry of covariant derivative, and with additivity of the "closer of quadrilaterals" $[\boldsymbol{u}, \boldsymbol{v}]$:

$$\boldsymbol{\nabla}_{\boldsymbol{u}+\boldsymbol{n}}\boldsymbol{v} = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} + \boldsymbol{\nabla}_{\boldsymbol{n}}\boldsymbol{v}.$$

(See exercise 10.1.) This can be inferred, alternatively, from the equivalence principle: in a local inertial frame, as in special relativity or Newtonian theory, the change in $\boldsymbol{v}$ along $\boldsymbol{u} + \boldsymbol{n}$ should equal the sum of the changes along $\boldsymbol{u}$ and along $\boldsymbol{n}$.

3. Consequence of above: choose $\boldsymbol{n}$ to be a multiple of $\boldsymbol{u}$; thereby conclude

$$\boldsymbol{\nabla}_{a\boldsymbol{u}}\boldsymbol{v} = a\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}.$$

Geodesics as defined by parallel transport or covariant differentiation

F. The "Schild's ladder" construction process for parallel transport (beginning of this box), applied to the tangent vector of a geodesic (exercise 10.6) guarantees: *a geodesic parallel transports its own tangent vector along itself.* Translated into covariant-derivative language:

$$\begin{pmatrix} \boldsymbol{u} = d/d\lambda \text{ is a tangent} \\ \text{vector to a curve, and} \\ \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \text{the curve is} \\ \text{a geodesic} \end{pmatrix}.$$

Thus closes the circle: geodesic to parallel transport to covariant derivative to geodesic.

2

Covariant derivative: basic properties

Symmetry: $\nabla_{\boldsymbol{u}}\boldsymbol{v} - \nabla_{\boldsymbol{v}}\boldsymbol{u} = [\boldsymbol{u}, \boldsymbol{v}]$ for any vector fields $\boldsymbol{u}$ and $\boldsymbol{v}$; (10.2a)

Chain rule: $\nabla_{\boldsymbol{u}}(f\boldsymbol{v}) = f\nabla_{\boldsymbol{u}}\boldsymbol{v} + \boldsymbol{v}\,\partial_{\boldsymbol{u}}f$ for any function f, vector field $\boldsymbol{v}$, and vector $\boldsymbol{u}$; (10.2b)

Additivity: $\nabla_{\boldsymbol{u}}(\boldsymbol{v} + \boldsymbol{w}) = \nabla_{\boldsymbol{u}}\boldsymbol{v} + \nabla_{\boldsymbol{u}}\boldsymbol{w}$ for any vector fields $\boldsymbol{v}$ and $\boldsymbol{w}$, and vector $\boldsymbol{u}$; (10.2c)

$\nabla_{a\boldsymbol{u}+b\boldsymbol{n}}\boldsymbol{v} = a\nabla_{\boldsymbol{u}}\boldsymbol{v} + b\nabla_{\boldsymbol{n}}\boldsymbol{v}$ for any vector field $\boldsymbol{v}$, vectors or vector fields $\boldsymbol{u}$ and $\boldsymbol{n}$, and numbers or functions a and b. (10.2d)

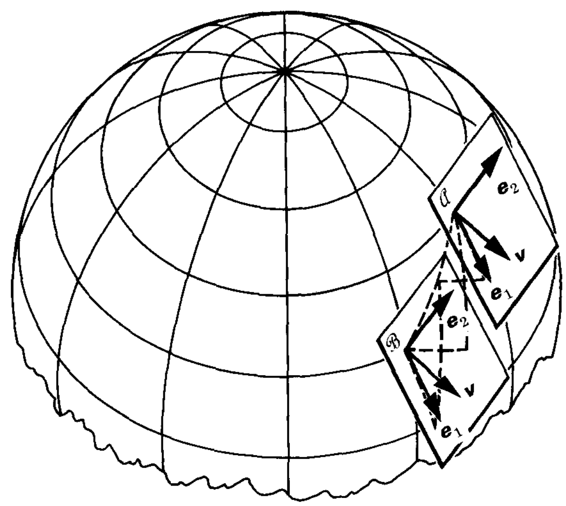


Figure 10.2.
The link between the tangent spaces at neighboring points, made possible by a parallel-transport law. Choose basis vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ at the event $\mathscr{A}$. Parallel transport them to a neighboring event $\mathscr{B}$. (Schild's ladder for transport of $\boldsymbol{e}_1$ is shown in the figure.) Then any other vector $\boldsymbol{v}$ that is parallel transported from $\mathscr{A}$ to $\mathscr{B}$ will have the same components at the two events (parallel transport cannot break the legs of a triangle; see Box 10.2):

[same numerically as at $\mathscr{A}$]

$$\boldsymbol{v} = v^1\boldsymbol{e}_1 + v^2\boldsymbol{e}_2 \text{ at } \mathscr{A} \Rightarrow \boldsymbol{v} = v^1\boldsymbol{e}_1 + v^2\boldsymbol{e}_2 \text{ at } \mathscr{B}.$$

[parallel transported from $\mathscr{A}$ to $\mathscr{B}$]

Thus, parallel transport provides a unique and complete link between the tangent space at $\mathscr{A}$ and the tangent space at $\mathscr{B}$. It identifies a unique vector at $\mathscr{B}$ with each vector at $\mathscr{A}$ in a way that preserves all algebraic relations. Similarly (see §10.3), it identifies a unique 1-form at $\mathscr{B}$ with each 1-form at $\mathscr{A}$, and a unique tensor at $\mathscr{B}$ with each tensor at $\mathscr{A}$, preserving all algebraic relations such as $\langle\boldsymbol{\sigma}, \boldsymbol{v}\rangle = 19.9$ and $\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{w}) = 37\,1$.

Actually, all this is true only in the limit when $\mathscr{A}$ and $\mathscr{B}$ are arbitrarily close to each other. When $\mathscr{A}$ and $\mathscr{B}$ are close but not arbitrarily close, the result of parallel transport is slightly different for different paths; so the link between the tangent spaces is slightly nonunique. But the differences decrease by a factor of 4 each time the affine-parameter distance between $\mathscr{A}$ and $\mathscr{B}$ is cut in half; see Chapter 11.

Any "rule" ∇, for producing new vector fields from old, that satisfies these four conditions, is called by differential geometers a *"symmetric covariant derivative."* Such a rule is not inherent in the more primitive concepts (Chapter 9) of curves, vectors, tensors, etc. In the arena of a spacetime laboratory, there are as many ways of defining a covariant derivative rule ∇ as there are of rearranging sources of the gravitational field. Different free-fall trajectories (geodesics) result from different distributions of masses.

Given the geodesics of spacetime, or of any other manifold, one can construct a unique corresponding covariant derivative by the Schild's ladder procedure of Box 10.2. Given any covariant derivative, one can discuss parallel transport via the equation

Equation for parallel transport

$$d\boldsymbol{v}/d\lambda \equiv \nabla_{\boldsymbol{u}}\boldsymbol{v} = 0 \Longleftrightarrow \text{the vector field } \boldsymbol{v} \text{ is parallel transported along the vector } \boldsymbol{u} = d/d\lambda; \tag{10.3}$$

and one can test whether any curve is a geodesic via

$$\begin{aligned} \nabla_{\boldsymbol{u}}\boldsymbol{u} = 0 &\Longleftrightarrow \text{the curve } \mathscr{P}(\lambda) \text{ with tangent vector } \boldsymbol{u} = d/d\lambda \\ &\quad\;\; \text{parallel transports its own tangent vector } \boldsymbol{u} \\ &\Longleftrightarrow \mathscr{P}(\lambda) \text{ is a geodesic.} \end{aligned} \tag{10.4}$$

Knowledge of all geodesics is equivalent to knowledge of covariant derivative

Thus a knowledge of all geodesics is completely equivalent to a knowledge of the covariant derivative.

Covariant derivative generalizes flat-space gradient

The covariant derivative ∇ generalizes to curved spacetime the flat-space gradient ∇. Like its flat-space cousin, it can be viewed as a machine for producing a number $\langle \boldsymbol{\sigma}, \nabla_{\boldsymbol{u}}\boldsymbol{v}\rangle$ out of a 1-form $\boldsymbol{\sigma}$, a vector $\boldsymbol{u}$, and a vector field $\boldsymbol{v}$. This machine viewpoint is explored in Box 10.3. Note there an important fact: despite its machine nature, ∇ is *not* a tensor; it is a nontensorial geometric object.

Action of covariant derivative on functions, 1-forms, and tensors

In curved as in flat spacetime, ∇ can be applied not only to vector fields, but also to functions, 1-form fields, and tensor fields. Its action on functions is defined in the obvious manner:

$$\nabla f \equiv \boldsymbol{d}f; \qquad \nabla_{\boldsymbol{u}} f \equiv \partial_{\boldsymbol{u}} f \equiv \boldsymbol{u}[f] \equiv \langle \boldsymbol{d}f, \boldsymbol{u}\rangle. \tag{10.5}$$

Its action on 1-form fields and tensor fields is defined by the curved-space generalization of equation (3.39): $\nabla \boldsymbol{S}$ is a linear machine for calculating the change in output of $\boldsymbol{S}$, from point to point, when "constant" (i.e., parallel transported) vectors are inserted into its slots. Example: the gradient of a $\binom{0}{1}$ tensor, i.e., of a 1-form field $\boldsymbol{\sigma}$. Pick an event $\mathscr{P}_0$; pick two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in the tangent space at $\mathscr{P}_0$; construct from $\boldsymbol{v}$ a "constant" vector field $\boldsymbol{v}(\mathscr{P})$ by parallel transport along the direction of $\boldsymbol{u}$, $\nabla_{\boldsymbol{u}}\boldsymbol{v} = 0$. Then $\nabla\boldsymbol{\sigma}$ is a $\binom{0}{2}$ tensor, and $\nabla_{\boldsymbol{u}}\boldsymbol{\sigma}$ is a $\binom{0}{1}$ tensor defined at $\mathscr{P}_0$ by

$$\nabla\boldsymbol{\sigma}(\boldsymbol{v}, \boldsymbol{u}) \equiv \langle \nabla_{\boldsymbol{u}}\boldsymbol{\sigma}, \boldsymbol{v}\rangle \equiv \nabla_{\boldsymbol{u}}(\langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle) \equiv \frac{d}{d\lambda}\langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle, \tag{10.6}$$

where $\boldsymbol{u} = d/d\lambda$. This defines $\nabla\boldsymbol{\sigma}$ and $\nabla_{\boldsymbol{u}}\boldsymbol{\sigma}$, because it states their output for any

(continued on page 257)

Box 10.3 COVARIANT DERIVATIVE VIEWED AS A MACHINE; CONNECTION COEFFICIENTS AS ITS COMPONENTS

A. The Machine View

1. The covariant derivative operator $\boldsymbol{\nabla}$, like most other geometric objects, can be regarded as a machine with slots. There is one such machine at each event $\mathcal{P}_0$ in spacetime. In brief, the machine interpretation of $\boldsymbol{\nabla}$ at $\mathcal{P}_0$ says

$$\boldsymbol{\nabla}(\boldsymbol{\sigma},\ \boldsymbol{v}(\mathcal{P}),\ \boldsymbol{u}) \equiv \langle \boldsymbol{\sigma}, \underbrace{\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v}} \rangle.$$

First slot; insert a 1-form $\boldsymbol{\sigma}$ that resides in the tangent space at $\mathcal{P}_0$

Second slot; insert a vector field $\boldsymbol{v}(\mathcal{P})$ defined on a neighborhood of $\mathcal{P}_0$

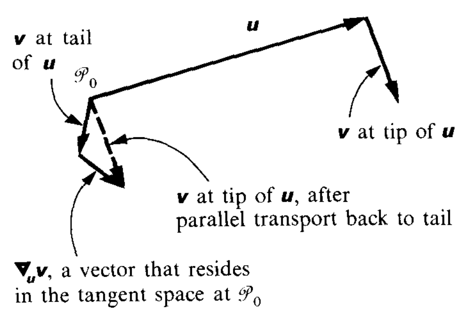
Third slot; insert a vector $\boldsymbol{u}$ that resides in the tangent space at $\mathcal{P}_0$

A new vector: "the covariant derivative of the vector field $\boldsymbol{v}$ with respect to $\boldsymbol{u}$."

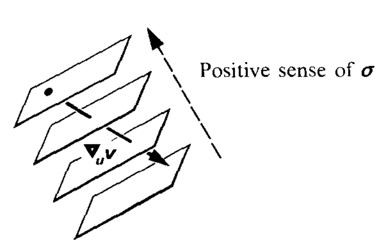
[Note: this slot notation for $\boldsymbol{\nabla}$ serves no useful purpose except to emphasize the "machine"-nature of $\boldsymbol{\nabla}$. This box is the only place it will be used.]

2. Geometrically, the output of the machine, $\langle \boldsymbol{\sigma}, \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v} \rangle$, is obtained as follows:

(a) Calculate the rate of change of $\boldsymbol{v}$, $\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v}$, along the vector $\boldsymbol{u}$; when $\boldsymbol{u}$ and $\boldsymbol{v}$ are infinitesimally small, the calculation can be represented pictorially:



(b) Count how many surfaces of the 1-form $\boldsymbol{\sigma}$ are pierced by the vector $\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v}$ (piercing occurs in tangent space at $\mathcal{P}_0$)



$$\langle \boldsymbol{\sigma}, \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v} \rangle = -2.8.$$

This number is the output of the machine ∇, when $\boldsymbol{\sigma}$, $\boldsymbol{v}(\mathscr{P})$ and $\boldsymbol{u}$ are inserted into its slots.

3. Another, equivalent, statement of covariant derivative as a machine. Leave first slot empty (no mention of any 1-form $\boldsymbol{\sigma}$); get a new vector field from original vector field $\boldsymbol{v}$:

$$\nabla(\underbrace{\ \ }_{\text{empty}}, \boldsymbol{v}(\mathscr{P}), \boldsymbol{u}) \equiv \nabla_{\boldsymbol{u}}\boldsymbol{v}$$

$$= \text{"covariant derivative of vector field } \boldsymbol{v} \text{ along vector } \boldsymbol{u}\text{."}$$

4. A third machine operation. Leave first and third slots empty (no mention of any 1-form $\boldsymbol{\sigma}$; no mention of any vector $\boldsymbol{u}$ along which to differentiate); get a $\binom{1}{1}$ tensor field from original vector field $\boldsymbol{v}$:

$$\nabla(\underbrace{\ \ }_{\text{empty}}, \boldsymbol{v}(\mathscr{P}), \underbrace{\ \ }_{\text{empty}}) \equiv \nabla\boldsymbol{v}$$

$$= \text{"covariant derivative" or "gradient" of vector field } \boldsymbol{v}.$$

This tensor field, $\nabla\boldsymbol{v}$, is the curved-space generalization of the flat-space $\nabla\boldsymbol{v}$ studied in §3.5. It has two slots (the two left empty in its definition). Its output for given input is

$$\nabla\boldsymbol{v}(\underbrace{\ \ }_{\text{empty}}, \boldsymbol{u}) = \nabla_{\boldsymbol{u}}\boldsymbol{v}$$

$$\nabla\boldsymbol{v}(\boldsymbol{\sigma}, \boldsymbol{u}) = \langle\boldsymbol{\sigma}, \nabla_{\boldsymbol{u}}\boldsymbol{v}\rangle.$$

5. Summary of the quantities defined above:
   (a) ∇ is a *covariant derivative operator;* to get a number from it, insert $\boldsymbol{\sigma}$, $\boldsymbol{v}(\mathscr{P})$, and $\boldsymbol{u}$; the result is $\langle\boldsymbol{\sigma}, \nabla_{\boldsymbol{u}}\boldsymbol{v}\rangle$.
   (b) $\nabla\boldsymbol{v}$ is the *gradient of* $\boldsymbol{v}$; to get a number from it, insert $\boldsymbol{\sigma}$ and $\boldsymbol{u}$; the result is $\langle\boldsymbol{\sigma}, \nabla_{\boldsymbol{u}}\boldsymbol{v}\rangle$ [same as in (a)].
   (c) $\nabla_{\boldsymbol{u}}\boldsymbol{v}$ is the *covariant derivative of* $\boldsymbol{v}$ *along* $\boldsymbol{u}$; to get a number from it, insert $\boldsymbol{\sigma}$; the result is $\langle\boldsymbol{\sigma}, \nabla_{\boldsymbol{u}}\boldsymbol{v}\rangle$ [same as in (a) and (b)].

B. How ∇ Differs from a Tensor

The machine ∇ differs from a tensor in two ways. (1) The middle slot of ∇ will not accept a vector; it demands a vector field—the vector field that is to be differentiated. (2) ∇ is not a linear machine (whereas a tensor must be linear!):

Box 10.3 (*continued*)

$$\begin{aligned}\nabla(a\sigma, f(\mathscr{P})\boldsymbol{v}(\mathscr{P}), b\boldsymbol{u}) &\equiv \langle a\sigma, \nabla_{b\boldsymbol{u}} f\boldsymbol{v}\rangle \\ &= abf\langle\sigma, \nabla_{\boldsymbol{u}}\boldsymbol{v}\rangle + \underbrace{ab\langle\sigma, \boldsymbol{v}\rangle\, \nabla_{\boldsymbol{u}} f}\end{aligned}$$

[this would be absent if ∇ were a linear machine.]

C. The "Connection Coefficients" as Components of ∇

Given a tensor $\boldsymbol{S}$ of rank $\binom{1}{2}$, a basis of tangent vectors $\{\boldsymbol{e}_\alpha\}$ at the event $\mathscr{P}_0$ where $\boldsymbol{S}$ resides, and the dual basis of 1-forms $\{\boldsymbol{\omega}^\alpha\}$, one defines the components of $\boldsymbol{S}$ by

$$S^\alpha{}_{\beta\gamma} \equiv \boldsymbol{S}(\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta, \boldsymbol{e}_\gamma).$$

One defines the components of ∇ similarly, except that for ∇ one needs not only a basis $\{\boldsymbol{e}_\alpha\}$ at the event $\mathscr{P}_0$, but also a basis $\{\boldsymbol{e}_\alpha(\mathscr{P})\}$ at each event $\mathscr{P}$ in its neighborhood:

$$\begin{aligned}\Gamma^\alpha{}_{\beta\gamma} &\equiv \text{components of } \nabla = \nabla(\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta(\mathscr{P}), \boldsymbol{e}_\gamma) \\ &\equiv \langle\boldsymbol{\omega}^\alpha, \nabla_{\boldsymbol{e}_\gamma}\boldsymbol{e}_\beta\rangle \\ &\simeq \begin{pmatrix}\text{“}\alpha\text{-component of change in basis vector } \boldsymbol{e}_\beta, \text{ when} \\ \text{in evaluating } \boldsymbol{e}_\beta \text{ one moves from tail to tip of } \boldsymbol{e}_\gamma\text{”}\end{pmatrix}.\end{aligned}$$

These components of ∇ are called the "*connection coefficients*" of the basis $\{\boldsymbol{e}_\alpha\}$. They are the "coordinate representation" of the covariant derivative operator ∇.

The covariant derivative operator ∇ and the connection coefficients $\Gamma^\alpha{}_{\mu\nu}$ provide different mathematical representations of the same geometric animal? Preposterous! The one animal runs from place to place and barks, or at least bites (takes difference, for example, between vector fields at one place and at a nearby place). The other animal, endowed with forty faces (see exercise 10.9) sits quietly at one spot. It would be difficult for two animals to look more different. Yet they do the same jobs in any world compatible with the equivalence principle: (1) they summarize the properties of all geodesics that go through the point in question; and, so doing, (2) they provide a physical means (parallel transport) to compare the values of vector fields and tensor fields at two neighboring events.

given input vectors $\boldsymbol{v}$ and $\boldsymbol{u}$. If $\boldsymbol{v}(\mathscr{P})$ is not constrained to be "constant" along $\boldsymbol{u} = d/d\lambda$, then $(d/d\lambda)\,\langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle$ has contributions from both the change in $\boldsymbol{v}$ and the change in $\boldsymbol{\sigma}$:

$$\frac{d}{d\lambda}\langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle \equiv \boldsymbol{\nabla}_{\boldsymbol{u}}\langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle = \langle \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\sigma}, \boldsymbol{v}\rangle + \langle \boldsymbol{\sigma}, \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}\rangle \tag{10.7}$$

(see exercise 10.3).

Similarly, if $\boldsymbol{S}$ is a $\binom{1}{2}$ tensor field, then its gradient $\boldsymbol{\nabla S}$ is a $\binom{1}{3}$ tensor field defined as follows. Pick an event $\mathscr{P}_0$; pick three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, and a 1-form $\boldsymbol{\sigma}$ in the tangent space at $\mathscr{P}_0$; turn $\boldsymbol{v}$, $\boldsymbol{w}$, and $\boldsymbol{\sigma}$ into "constant" vector fields and a "constant" 1-form field near $\mathscr{P}_0$ by means of parallel transport ($\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w} = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\sigma} = 0$ at $\mathscr{P}_0$); then define

$$\begin{aligned}\boldsymbol{\nabla S}(\boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u}) &\equiv (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{S})(\boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{w}) \equiv \boldsymbol{\nabla}_{\boldsymbol{u}}[\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{w})]\\ &= \partial_{\boldsymbol{u}}[\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{w})].\end{aligned} \tag{10.8}$$

EXERCISES

Exercise 10.1. ADDITIVITY OF COVARIANT DIFFERENTIATION

Show that the commutator ("closer of quadrilaterals") is additive:

$$[\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{w}] = [\boldsymbol{u}, \boldsymbol{v}] + [\boldsymbol{u}, \boldsymbol{w}]; \qquad [\boldsymbol{u} + \boldsymbol{n}, \boldsymbol{v}] = [\boldsymbol{u}, \boldsymbol{v}] + [\boldsymbol{n}, \boldsymbol{v}].$$

Use this result, the additivity condition $\boldsymbol{\nabla}_{\boldsymbol{u}}(\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} + \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}$, and symmetry of the covariant derivative, $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} - \boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{u} = [\boldsymbol{u}, \boldsymbol{v}]$, to prove that

$$\boldsymbol{\nabla}_{\boldsymbol{u}+\boldsymbol{n}}\boldsymbol{v} = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} + \boldsymbol{\nabla}_{\boldsymbol{n}}\boldsymbol{v}.$$

Exercise 10.2. CHAIN RULE FOR COVARIANT DIFFERENTIATION

Use pictures, and the "take-the-difference-and-take-the-limit" definition of $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}$ (Box 10.2) to show that

$$\boldsymbol{\nabla}_{\boldsymbol{u}}(f\boldsymbol{v}) = f\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} + \boldsymbol{v}\,\partial_{\boldsymbol{u}}[f]. \tag{10.9}$$

Exercise 10.3. ANOTHER CHAIN RULE

Derive equation (10.7), using the "take-the-difference-and-take-the-limit" definitions of derivatives. *Hint:* Before taking the differences, parallel transport $\boldsymbol{\sigma}[\mathscr{P}(\lambda)]$ and $\boldsymbol{v}[\mathscr{P}(\lambda)]$ back from $\mathscr{P}(\lambda)$ to $\mathscr{P}(0)$.

Exercise 10.4. STILL ANOTHER CHAIN RULE

Show that, as in flat spacetime, so also in curved spacetime,

$$\boldsymbol{\nabla}_{\boldsymbol{u}}(\boldsymbol{v} \otimes \boldsymbol{w}) = (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}) \otimes \boldsymbol{w} + \boldsymbol{v} \otimes (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}). \tag{10.10}$$

Write down the more familiar component version of this equation in flat spacetime.

Solution to first part of exercise: Choose 1-forms $\boldsymbol{\sigma}$ and $\boldsymbol{\rho}$ at the event $\mathscr{P}_0$ in question, and extend them along the vector $\boldsymbol{u} = d/d\lambda$ by parallel transport, $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\rho} = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\sigma} = 0$. Then

$$[\boldsymbol{\nabla}_{\boldsymbol{u}}(\boldsymbol{v} \otimes \boldsymbol{w})](\boldsymbol{\rho}, \boldsymbol{\sigma}) = \frac{d}{d\lambda}[(\boldsymbol{v} \otimes \boldsymbol{w})(\boldsymbol{\rho}, \boldsymbol{\sigma})] \qquad \text{(def of } \boldsymbol{\nabla}_{\boldsymbol{u}} \text{ on a tensor)}$$

$$= \frac{d}{d\lambda}[\langle \boldsymbol{\rho}, \boldsymbol{v}\rangle\langle \boldsymbol{\sigma}, \boldsymbol{w}\rangle] \qquad \text{(def of tensor product "}\otimes\text{")}$$

$$= \frac{d\langle \boldsymbol{\rho}, \boldsymbol{v}\rangle}{d\lambda}\langle \boldsymbol{\sigma}, \boldsymbol{w}\rangle + \langle \boldsymbol{\rho}, \boldsymbol{v}\rangle \frac{d\langle \boldsymbol{\sigma}, \boldsymbol{w}\rangle}{d\lambda} \qquad \text{(chain rule for derivatives)}$$

$$= \langle \boldsymbol{\rho}, \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}\rangle\langle \boldsymbol{\sigma}, \boldsymbol{w}\rangle + \langle \boldsymbol{\rho}, \boldsymbol{v}\rangle\langle \boldsymbol{\sigma}, \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}\rangle$$
$$\text{(by equation 10.7 with } \boldsymbol{\rho}, \boldsymbol{\sigma} \text{ const)}$$

$$= [(\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}) \otimes \boldsymbol{w}](\boldsymbol{\rho}, \boldsymbol{\sigma}) + [\boldsymbol{v} \otimes (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w})](\boldsymbol{\rho}, \boldsymbol{\sigma})$$
$$\text{(def of tensor product "}\otimes\text{").}$$

Exercise 10.5. ONE MORE CHAIN RULE

Show, using techniques similar to those in exercise 10.4, that

$$\boldsymbol{\nabla}_{\boldsymbol{u}}(\boldsymbol{\sigma} \otimes \boldsymbol{\rho} \otimes \boldsymbol{v}) = (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\sigma}) \otimes \boldsymbol{\rho} \otimes \boldsymbol{v} + \boldsymbol{\sigma} \otimes (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\rho}) \otimes \boldsymbol{v} + \boldsymbol{\sigma} \otimes \boldsymbol{\rho} \otimes (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}). \qquad (10.11)$$

Exercise 10.6. GEODESIC EQUATION

Use the "Schild's ladder" construction process for parallel transport (beginning of Box 10.2) to show that a geodesic parallel transports its own tangent vector along itself (end of Box 10.2).

§10.4. PARALLEL TRANSPORT AND COVARIANT DERIVATIVE: COMPONENT APPROACH

The pictorial approach motivates the mathematics; the abstract approach makes the pictorial ideas precise; but usually one must use the component approach in order to actually do complex calculations.

To work with components, one needs a set of basis vectors $\{\boldsymbol{e}_\alpha\}$ and the dual set of basis 1-forms $\{\boldsymbol{\omega}^\alpha\}$. In flat spacetime a single such basis suffices; all events can use the same Lorentz basis. Not so in curved spacetime! There each event has its own tangent space, and each tangent space requires a basis of its own. As one travels from event to event, comparing their bases via parallel transport, one sees the bases twist and turn. They must do so. In no other way can they accommodate themselves to the curvature of spacetime. Bases at points $\mathscr{P}_0$ and $\mathscr{P}_1$, which are the same when compared by parallel transport along one curve, must differ when compared along another curve (see "Curvature"; Chapter 11).

To quantify the twisting and turning of a "field" of basis vectors $\{\boldsymbol{e}_\alpha(\mathscr{P})\}$ and forms $\{\boldsymbol{\omega}^\alpha(\mathscr{P})\}$, use the covariant derivative. Examine the changes in vector fields along a basis vector $\boldsymbol{e}_\beta$, abbreviating

$$\boldsymbol{\nabla}_{\boldsymbol{e}_\beta} \equiv \boldsymbol{\nabla}_\beta \qquad \text{(def of } \boldsymbol{\nabla}_\beta\text{);} \qquad (10.12)$$

and especially examine the rate of change of some basis vector: $\boldsymbol{\nabla}_\beta \boldsymbol{e}_\alpha$. This rate of change is itself a vector, so it can be expanded in terms of the basis:

$$\nabla_{\underbrace{\beta}}\boldsymbol{e}_{\underbrace{\alpha}} = \boldsymbol{e}_\mu \Gamma^\mu{}_{\underbrace{\alpha\beta}} \qquad (\text{def of } \Gamma^\mu{}_{\alpha\beta}); \tag{10.13}$$

note reversal of order of α and β!

Connection coefficients defined

and the resultant *"connection coefficients"* $\Gamma^\mu{}_{\alpha\beta}$ can be calculated by projection on the basis 1-forms:

$$\langle \boldsymbol{\omega}^\mu, \nabla_\beta \boldsymbol{e}_\alpha \rangle = \Gamma^\mu{}_{\alpha\beta}. \tag{10.14}$$

(See exercise 10.7; also Box 10.3.) Because the basis 1-forms are "locked into" the basis vectors ($\langle \boldsymbol{\omega}^\nu, \boldsymbol{e}_\alpha \rangle = \delta^\nu{}_\alpha$), these same connection coefficients $\Gamma^\nu{}_{\alpha\beta}$ tell how the 1-form basis changes from point to point:

$$\nabla_\beta \boldsymbol{\omega}^\nu = -\Gamma^\nu{}_{\alpha\beta}\boldsymbol{\omega}^\alpha, \tag{10.15}$$

$$\langle \nabla_\beta \boldsymbol{\omega}^\nu, \boldsymbol{e}_\alpha \rangle = -\Gamma^\nu{}_{\alpha\beta}. \tag{10.16}$$

(See exercise 10.8.)

The connection coefficients do even more. They allow one to calculate the components of the gradient of an arbitrary tensor $\boldsymbol{S}$. In a Lorentz frame of flat spacetime, the components of $\nabla\boldsymbol{S}$ are obtained by letting the basis vectors $\boldsymbol{e}_\alpha = \partial\mathcal{P}/\partial x^\alpha = \partial/\partial x^\alpha$ act on the components of $\boldsymbol{S}$. Thus for a $\binom{1}{2}$ tensor field $\boldsymbol{S}$ one finds that

$$\nabla\boldsymbol{S} \text{ has components } S^\alpha{}_{\beta\gamma,\delta} = \frac{\partial}{\partial x^\delta}[S^\alpha{}_{\beta\gamma}].$$

Not so in curved spacetime, or even in a non-Lorentz basis in flat spacetime. There the basis vectors turn, twist, expand, and contract, so even if $\boldsymbol{S}$ were constant ($\nabla\boldsymbol{S} = 0$), its components on the twisting basis vectors would vary. The connection coefficients, properly applied, will compensate for this twisting and turning. As one learns in exercise 10.10, the components of $\nabla\boldsymbol{S}$, called $S^\alpha{}_{\beta\gamma;\delta}$ so that

Components of the gradient of a tensor field

$$\nabla\boldsymbol{S} = S^\alpha{}_{\beta\gamma;\delta}\boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \otimes \boldsymbol{\omega}^\gamma \otimes \boldsymbol{\omega}^\delta, \tag{10.17}$$

can be calculated from those of $\boldsymbol{S}$ by the usual flat-space method, plus a correction applied to each index (i.e., to each basis vector):

["+" when correcting "up" indices] [interchange and sum on index being corrected]

[differentiating index]

$$S^\alpha{}_{\beta\gamma;\delta} = S^\alpha{}_{\beta\gamma,\delta} + S^\mu{}_{\beta\gamma}\Gamma^\alpha{}_{\mu\delta} - S^\alpha{}_{\mu\gamma}\Gamma^\mu{}_{\beta\delta} - S^\alpha{}_{\beta\mu}\Gamma^\mu{}_{\gamma\delta}. \tag{10.18}$$

["−" when correcting "down" indices]

[interchange and sum on index being corrected] [differentiating index]

Here

$$S^\alpha{}_{\beta\gamma,\delta} \equiv \boldsymbol{e}_\delta[S^\alpha{}_{\beta\gamma}] \equiv \partial_{\boldsymbol{e}_\delta} S^\alpha{}_{\beta\gamma}. \tag{10.19}$$

Equation (10.18) looks complicated; but it is really very simple, once the pattern has been grasped.

Components of the covariant derivative of a tensor field

Just as one uses special notation, $S^{\alpha}{}_{\beta\gamma;\delta}$, for the components of $\boldsymbol{\nabla S}$, so one introduces special notation, $DS^{\alpha}{}_{\beta\gamma}/d\lambda$, for components of the covariant derivative $\boldsymbol{\nabla_u S}$ along $\boldsymbol{u} = d/d\lambda$:

$$\boldsymbol{\nabla_u S} = (DS^{\alpha}{}_{\beta\gamma}/d\lambda)\boldsymbol{e}_{\alpha} \otimes \boldsymbol{\omega}^{\beta} \otimes \boldsymbol{\omega}^{\gamma}; \tag{10.20}$$

$$\frac{DS^{\alpha}{}_{\beta\gamma}}{d\lambda} = S^{\alpha}{}_{\beta\gamma;\delta}u^{\delta} = (S^{\alpha}{}_{\beta\gamma,\delta} + \text{correction terms})u^{\delta}.$$

Since for any f

$$f_{,\delta}u^{\delta} = \partial_{\boldsymbol{u}} f = df/d\lambda$$

this reduces to

$$\frac{DS^{\alpha}{}_{\beta\gamma}}{d\lambda} = \frac{dS^{\alpha}{}_{\beta\gamma}}{d\lambda} + S^{\mu}{}_{\beta\gamma}\Gamma^{\alpha}{}_{\mu\delta}u^{\delta} - S^{\alpha}{}_{\mu\gamma}\Gamma^{\mu}{}_{\beta\delta}u^{\delta} - S^{\alpha}{}_{\beta\mu}\Gamma^{\mu}{}_{\gamma\delta}u^{\delta}. \tag{10.21}$$

Chain rule for gradient

The power of the component approach shows up clearly when one discusses chain rules for covariant derivatives. The multitude of abstract-approach chain rules (equations 10.2b, 10.7, 10.10, 10.11) all boil down into a single rule for components: *The gradient operation ";" obeys the standard partial-differentiation chain rule of ordinary calculus.* Example:

$$(fv^{\alpha})_{;\mu} = \underbrace{f_{;\mu}}_{[= f_{,\mu} \text{ because } f \text{ has no indices to correct}]} v^{\alpha} + fv^{\alpha}{}_{;\mu} \tag{10.22a}$$

(contract this with u^{μ} to get chain rule 10.2b). Another example:

$$\underbrace{(\sigma_{\alpha}v^{\alpha})_{;\mu}}_{[= (\sigma_{\alpha}v^{\alpha})_{,\mu} \text{ because } \sigma_{\alpha}v^{\alpha} \text{ has no free indices to correct}]} = \sigma_{\alpha;\mu}v^{\alpha} + \sigma_{\alpha}v^{\alpha}{}_{;\mu} \tag{10.22b}$$

(contract this with u^{μ} to get chain rule 10.7). Another example:

$$(\sigma_{\alpha}\rho_{\beta}v^{\gamma})_{;\mu} = \sigma_{\alpha;\mu}\rho_{\beta}v^{\gamma} + \sigma_{\alpha}\rho_{\beta;\mu}v^{\gamma} + \sigma_{\alpha}\rho_{\beta}v^{\gamma}{}_{;\mu} \tag{10.22c}$$

(contract this with u^{μ} to get chain rule 10.11). Another example: see Exercise (10.12) below.

EXERCISES

Exercise 10.7. COMPUTATION OF CONNECTION COEFFICIENTS

Derive equation (10.14) for $\Gamma^{\mu}{}_{\alpha\beta}$ from equation (10.13).

Exercise 10.8. CONNECTION FOR 1-FORM BASIS

Derive equations (10.15) and (10.16), which relate $\boldsymbol{\nabla}_{\beta}\boldsymbol{\omega}^{\nu}$ *to* $\Gamma^{\nu}{}_{\alpha\beta}$, from equation (10.14). *Hint:* use equation (10.7).

Exercise 10.9. SYMMETRY OF CONNECTION COEFFICIENTS

Show that the symmetry of spacetime's covariant derivative (equation 10.2a) is equivalent to the following symmetry condition on the connection coefficients:

$$\text{(antisymmetric part of } \Gamma^{\mu}{}_{\alpha\beta}) \equiv \frac{1}{2}(\Gamma^{\mu}{}_{\alpha\beta} - \Gamma^{\mu}{}_{\beta\alpha})$$

$$\equiv \Gamma^{\mu}{}_{[\alpha\beta]} = -\frac{1}{2}\langle \boldsymbol{\omega}^{\mu}, \underbrace{[\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta]}_{\text{[commutator of basis vectors]}}\rangle \equiv -\frac{1}{2}c_{\alpha\beta}{}^{\mu}. \qquad (10.23)$$

As a special case, $\Gamma^{\mu}{}_{\alpha\beta}$ *is symmetric in* α *and* β *when a coordinate basis* ($\boldsymbol{e}_\alpha = \partial/\partial x^\alpha$) *is used.* Show that in a coordinate basis this symmetry reduces the number of independent connection coefficients at each event from $4 \times 4 \times 4 = 64$ to $4 \times 10 = 40$.

Exercise 10.10. COMPONENTS OF GRADIENT

Derive equation (10.18) for the components of the gradient, $S^{\alpha}{}_{\beta\gamma;\delta}$. *Hint:* Expand $\boldsymbol{S}$ in terms of the given basis, and then evaluate the righthand side of

$$\nabla_{\boldsymbol{u}}\boldsymbol{S} = \nabla_{\boldsymbol{u}}(S^{\alpha}{}_{\beta\gamma}\boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^{\beta} \otimes \boldsymbol{\omega}^{\gamma}),$$

for an arbitrary vector $\boldsymbol{u}$. Use the chain rules (10.2b) and (10.11). By comparing the result with

$$\nabla_{\boldsymbol{u}}\boldsymbol{S} = S^{\alpha}{}_{\beta\gamma;\delta}u^{\delta}\boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^{\beta} \otimes \boldsymbol{\omega}^{\gamma},$$

read off the components $S^{\alpha}{}_{\beta\gamma;\delta}$.

Exercise 10.11. DIVERGENCE

Let $\boldsymbol{T}$ be a $\binom{2}{0}$ tensor field, and define the divergence on its second slot by the same process as in flat spacetime: $\nabla \cdot \boldsymbol{T}$ = contraction of $\nabla\boldsymbol{T}$; i.e.,

$$(\nabla \cdot \boldsymbol{T})^{\alpha} = T^{\alpha\beta}{}_{;\beta}. \qquad (10.24)$$

Write the components $T^{\alpha\beta}{}_{;\beta}$ in terms of $T^{\alpha\beta}{}_{,\beta}$ plus correction terms for each of the two indices of $\boldsymbol{T}$.

[*Answer:*

$$T^{\alpha\beta}{}_{;\beta} = T^{\alpha\beta}{}_{,\beta} + \Gamma^{\alpha}{}_{\mu\beta}T^{\mu\beta} + \Gamma^{\beta}{}_{\mu\beta}T^{\alpha\mu}.]$$

Exercise 10.12. VERIFICATION OF CHAIN RULE

Let $S^{\alpha\beta}{}_{\gamma}$ be components of a $\binom{2}{1}$ tensor field, and $M_{\beta}{}^{\gamma}$ be components of a $\binom{1}{1}$ tensor field. By contracting these tensor fields, one obtains a vector field $S^{\alpha\beta}{}_{\gamma}M_{\beta}{}^{\gamma}$. The chain rule for the divergence of this vector field reads

$$(S^{\alpha\beta}{}_{\gamma}M_{\beta}{}^{\gamma})_{;\alpha} = S^{\alpha\beta}{}_{\gamma;\alpha}M_{\beta}{}^{\gamma} + S^{\alpha\beta}{}_{\gamma}M_{\beta}{}^{\gamma}{}_{;\alpha}.$$

Verify the validity of this chain rule by expressing both sides of the equation in terms of directional derivatives $(\partial_{\boldsymbol{e}_\alpha})$ plus connection-coefficient corrections. *Hint:* the left side becomes

$$(S^{\alpha\beta}{}_{\gamma}M_{\beta}{}^{\gamma})_{;\alpha} = \underbrace{(S^{\alpha\beta}{}_{\gamma}M_{\beta}{}^{\gamma})_{,\alpha}}_{\left[\substack{S^{\alpha\beta}{}_{\gamma,\alpha}M_{\beta}{}^{\gamma} + S^{\alpha\beta}{}_{\gamma}M_{\beta}{}^{\gamma}{}_{,\alpha} \\ \text{by chain rule for directional derivative}}\right]} + \Gamma^{\alpha}{}_{\mu\alpha}(S^{\mu\beta}{}_{\gamma}M_{\beta}{}^{\gamma}).$$

The right side has many more correction terms (three on $S^{\alpha\beta}{}_{\gamma;\alpha}$; two on $M_\beta{}^\gamma{}_{;\alpha}$), but they must cancel against each other, leaving only one.

Exercise 10.13. TRANSFORMATION LAW FOR CONNECTION COEFFICIENTS

Let $\{\boldsymbol{e}_\alpha\}$ and $\{\boldsymbol{e}_{\mu'}\}$ be two different fields of basis vectors related by the transformation law

$$\boldsymbol{e}_{\mu'}(\mathscr{P}) = L^\alpha{}_{\mu'}(\mathscr{P})\boldsymbol{e}_\alpha(\mathscr{P}). \tag{10.25}$$

Show that the corresponding connection coefficients are related by

$$\Gamma^{\alpha'}{}_{\beta'\gamma'} = \underbrace{L^{\alpha'}{}_\rho L^\mu{}_{\beta'} L^\nu{}_{\gamma'} \Gamma^\rho{}_{\mu\nu}}_{\substack{\text{standard transformation law}\\ \text{for components of a tensor}}} + L^{\alpha'}{}_\mu L^\mu{}_{\beta',\gamma'} \tag{10.26}$$

Exercise 10.14. POLAR COORDINATES IN FLAT 2-DIMENSIONAL SPACE

On a sheet of paper draw an (r, ϕ) polar coordinate system. At neighboring points, draw the basis vectors $\boldsymbol{e}_{\hat{r}} = \partial/\partial r$ and $\boldsymbol{e}_{\hat{\phi}} \equiv r^{-1}\,\partial/\partial\phi$. (a) Use this picture, and Euclid's version of parallel transport, to justify the relations

$$\nabla_{\hat{r}}\boldsymbol{e}_{\hat{r}} = 0, \qquad \nabla_{\hat{r}}\boldsymbol{e}_{\hat{\phi}} = 0, \qquad \nabla_{\hat{\phi}}\boldsymbol{e}_{\hat{r}} = r^{-1}\boldsymbol{e}_{\hat{\phi}}, \qquad \nabla_{\hat{\phi}}\boldsymbol{e}_{\hat{\phi}} = -r^{-1}\boldsymbol{e}_{\hat{r}}.$$

(b) From these relations write down the connection coefficients. (c) Let $\boldsymbol{A} = A^{\hat{r}}\boldsymbol{e}_{\hat{r}} + A^{\hat{\phi}}\boldsymbol{e}_{\hat{\phi}}$ be a vector field. Show that its divergence, $\nabla\cdot\boldsymbol{A} = A^{\hat{\alpha}}{}_{;\hat{\alpha}} = A^{\hat{\alpha}}{}_{,\hat{\alpha}} + \Gamma^{\hat{\alpha}}{}_{\hat{\mu}\hat{\alpha}}A^{\hat{\mu}}$, can be calculated using the formula

$$\nabla\cdot\boldsymbol{A} = \frac{1}{r}\frac{\partial A^{\hat{\phi}}}{\partial\phi} + \frac{1}{r}\frac{\partial(rA^{\hat{r}})}{\partial r}$$

(which should be familiar to most readers).

§10.5. GEODESIC EQUATION

Geodesic equation: abstract version

Geodesics—the parametrized paths of freely falling particles—were the starting point of this chapter. From them parallel transport was constructed (Schild's ladder; Box 10.2); and parallel transport in turn produced the covariant derivative and its connection coefficients. Given the covariant derivative, one recovered the geodesics: they were the curves whose tangent vectors, $\boldsymbol{u} = d\mathscr{P}/d\lambda$, satisfy $\nabla_{\boldsymbol{u}}\boldsymbol{u} = 0$ ($\boldsymbol{u}$ is parallel transported along itself).

Let a coordinate system $\{x^\alpha(\mathscr{P})\}$ be given. Let it induce basis vectors $\boldsymbol{e}_\alpha = \partial/\partial x^\alpha$ into the tangent space at each event. Let the connection coefficients $\Gamma^\alpha{}_{\beta\gamma}$ for this "coordinate basis" be given. Then the component version of the "*geodesic equation*" $\nabla_{\boldsymbol{u}}\boldsymbol{u} = 0$ becomes a differential equation for the geodesic $x^\alpha(\lambda)$:

(1) $\boldsymbol{u} = \dfrac{d}{d\lambda} = \dfrac{dx^\alpha}{d\lambda}\dfrac{\partial}{\partial x^\alpha} \implies$ components of $\boldsymbol{u}$ are $u^\alpha = \dfrac{dx^\alpha}{d\lambda}$;

(2) then components of $\nabla_{\boldsymbol{u}}\boldsymbol{u} = 0$ are

$$0 = u^\alpha{}_{;\beta}u^\beta = (u^\alpha{}_{,\beta} + \Gamma^\alpha{}_{\gamma\beta}u^\gamma)u^\beta$$
$$= \frac{\partial}{\partial x^\beta}\left(\frac{dx^\alpha}{d\lambda}\right)\frac{dx^\beta}{d\lambda} + \Gamma^\alpha{}_{\gamma\beta}\frac{dx^\gamma}{d\lambda}\frac{dx^\beta}{d\lambda},$$

which reduces to the differential equation

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha{}_{\gamma\beta}\frac{dx^\gamma}{d\lambda}\frac{dx^\beta}{d\lambda} = 0. \tag{10.27}$$

Component version

This component version of the geodesic equation gives an analytic method ("translation" of Schild's ladder) for constructing the parallel transport law from a knowledge of the geodesics. Pick an event $\mathscr{P}_0$ and set up a coordinate system in its neighborhood. Watch many clock-carrying particles pass through (or arbitrarily close to) $\mathscr{P}_0$. For each particle read off the values of $d^2x^\alpha/d\lambda^2$ and $dx^\alpha/d\lambda$ at $\mathscr{P}_0$. Insert all the data for many particles into equation (10.27), and solve for the connection coefficients. Do not be disturbed that only the symmetric part of $\Gamma^\alpha{}_{\gamma\beta}$ is obtained thereby; the antisymmetric part, $\Gamma^\alpha{}_{[\gamma\beta]}$, vanishes identically in any coordinate frame! (See exercise 10.9.) Knowing $\Gamma^\alpha{}_{\gamma\beta}$, use them to parallel transport any desired vector along any desired curve through $\mathscr{P}_0$:

How to construct parallel transport law from knowledge of geodesics

$$\nabla_{\boldsymbol{u}}\boldsymbol{v} = 0 \quad \Longleftrightarrow \quad \frac{dv^\alpha}{d\lambda} + \Gamma^\alpha{}_{\gamma\beta}v^\gamma\frac{dx^\beta}{d\lambda} = 0. \tag{10.28}$$

EXERCISES

Exercise 10.15. COMPONENTS OF PARALLEL-TRANSPORT LAW

Show that equation (10.28) is the component version of the law for parallel transporting a vector $\boldsymbol{v}$ along the curve $\mathscr{P}(\lambda)$ with tangent vector $\boldsymbol{u} = d\mathscr{P}/d\lambda$.

Exercise 10.16. GEODESICS IN POLAR COORDINATES

In rectangular coordinates on a flat sheet of paper, Euclid's straight lines (geodesics) satisfy $d^2x/d\lambda^2 = d^2y/d\lambda^2 = 0$. Transform this geodesic equation into polar coordinates ($x = r\cos\phi$, $y = r\sin\phi$); and read off the resulting connection coefficients by comparison with equation (10.27). These are the connection coefficients for the coordinate basis ($\partial/\partial r$, $\partial/\partial\phi$). From them calculate the connection coefficients for the basis

$$\boldsymbol{e}_{\hat r} = \frac{\partial}{\partial r}, \qquad \boldsymbol{e}_{\hat\phi} = \frac{1}{r}\frac{\partial}{\partial\phi}.$$

The answer should agree with the answer to part (b) of Exercise 10.14. *Hint:* Use such relations as

$$\nabla_{\boldsymbol{e}_{\hat\phi}}\boldsymbol{e}_{\hat r} = \nabla_{(1/r)(\partial/\partial\phi)}(\partial/\partial r) = \frac{1}{r}\nabla_{(\partial/\partial\phi)}(\partial/\partial r).$$

2

Exercise 10.17. ROTATION GROUP: GEODESICS AND CONNECTION COEFFICIENTS

[Continuation of exercises 9.13 and 9.14.] In discussing the rotation group, one must make a clear distinction between the *Euclidean space* (coordinates x, y, z; basis vectors $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$) in which the rotation matrices act, and the *group manifold* $SO(3)$ (coordinates ψ, θ, ϕ; coordinate basis $\partial/\partial\psi$, $\partial/\partial\theta$, $\partial/\partial\phi$; basis of "generators" $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$), whose points $\mathscr{P}$ are rotation matrices.

(a) Pick a vector

$$\boldsymbol{n} = n^x\, \partial/\partial x + n^y\, \partial/\partial y + n^z\, \partial/\partial z$$

in Euclidean space. Show that

$$\mathscr{R}_{\boldsymbol{n}}(t) \equiv \exp[(n^x\mathscr{K}_1 + n^y\mathscr{K}_2 + n^z\mathscr{K}_3)t] \tag{10.29}$$

is a rotation matrix that rotates the axes of Euclidean space by an angle

$$t|\boldsymbol{n}| = t[(n^x)^2 + (n^y)^2 + (n^z)^2]^{1/2}$$

about the direction $\boldsymbol{n}$. ($\mathscr{K}_j$ are matrices defined in exercise 9.13.)

(b) In the group manifold $SO(3)$, pick a point (rotation matrix) $\mathscr{P}$, and pick a tangent vector $\boldsymbol{u} = u^\alpha \boldsymbol{e}_\alpha$ at $\mathscr{P}$. Let $\boldsymbol{u}$ be a vector in Euclidean space with the same components as $\boldsymbol{u}$ has in $SO(3)$:

$$\boldsymbol{u} = u^1\boldsymbol{e}_1 + u^2\boldsymbol{e}_2 + u^3\boldsymbol{e}_3; \qquad \boldsymbol{u} = u^1\, \partial/\partial x + u^2\, \partial/\partial y + u^3\, \partial/\partial z. \tag{10.30}$$

Show that $\boldsymbol{u}$ is the tangent vector (at $t = 0$) to the curve

$$\mathscr{C}(t) = \mathscr{R}_{\boldsymbol{u}}(t)\mathscr{P}. \tag{10.31}$$

The curve $\mathscr{C}(t)$ through the arbitrary point $\mathscr{P}$ with arbitrary tangent vector $\boldsymbol{u} = (d\mathscr{C}/dt)_{t=0}$ is a very special curve: every point on it differs from $\mathscr{P}$ by a rotation $\mathscr{R}_{\boldsymbol{u}}(t)$ about one and the same direction $\boldsymbol{u}$. No other curve in $SO(3)$ with "starting conditions" $\{\mathscr{P}, \boldsymbol{u}\}$ has such beautiful simplicity. Hence it is natural to decree that each such $\mathscr{C}(t)$ is a geodesic of the group manifold $SO(3)$. This decree adds new geometric structure to $SO(3)$; it converts $SO(3)$ from a differentiable manifold into something more special: an *affine manifold.*

One has no guarantee that an arbitrarily chosen family of curves in an arbitrary manifold *can* be decreed to be geodesics. Most families of curves simply do not possess the right geometric properties to function as geodesics. Most will lead to covariant derivatives that violate one or more of the fundamental conditions (10.2). To learn whether a given choice of geodesics is possible, one can try to derive connection coefficients $\Gamma^\alpha{}_{\beta\gamma}$ (for some given basis) corresponding to the chosen geodesics. If the derivation is successful, the choice of geodesics was a possible one. If the derivation produces inconsistencies, the chosen family of curves have the wrong geometric properties to function as geodesics.

(c) For the basis of generators $\{\boldsymbol{e}_\alpha\}$ derive connection coefficients corresponding to the chosen geodesics, $\mathscr{C}(t) = \mathscr{R}_{\boldsymbol{u}}(t)\mathscr{P}$, of $SO(3)$. *Hint:* show that the components $u^\alpha = \langle \boldsymbol{\omega}^\alpha, \boldsymbol{u}\rangle$ of the tangent $\boldsymbol{u} = d\mathscr{C}/dt$ to a given geodesic are independent of position $\mathscr{C}(t)$ along the geodesic. Then use the geodesic equation $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0$, expanded in the basis $\{\boldsymbol{e}_\alpha\}$, to calculate the symmetric part of the connection $\Gamma^\alpha{}_{(\beta\gamma)}$. Finally use equation (10.23) to calculate $\Gamma^\alpha{}_{[\beta\gamma]}$. [*Answer:*

$$\Gamma^\alpha{}_{\beta\gamma} = \frac{1}{2}\epsilon_{\alpha\beta\gamma}, \tag{10.32}$$

where $\epsilon_{\alpha\beta\gamma}$ is the completely antisymmetric symbol with $\epsilon_{123} = +1$. This answer is independent of location $\mathscr{P}$ in $SO(3)$!]

CHAPTER 11

GEODESIC DEVIATION AND SPACETIME CURVATURE

§11.1. CURVATURE, AT LAST!

This chapter is entirely Track 2. Chapters 9 and 10 are necessary preparation for it.
It will be needed as preparation for
(1) Chapters 12 and 13 (Newtonian gravity; Riemannian geometry),
(2) the second half of Chapter 14 (calculation of curvature), and
(3) the details, but not the message, of Chapter 15 (Bianchi identities).

Spacetime curvature manifests itself as gravitation, by means of the deviation of one geodesic from a nearby geodesic (relative acceleration of test particles).

Overview of chapter

Let the geodesics of spacetime be known. Then the covariant derivative $\boldsymbol{\nabla}$ and its connection coefficients $\Gamma^{\alpha}{}_{\beta\gamma}$ are also known. How, from this information, does one define, calculate, and understand geodesic deviation and spacetime curvature? The answer unfolds in this chapter, and is summarized in Box 11.1. To disclose the answer one must (1) define the "relative acceleration vector" $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{n}$, which measures the deviation of one geodesic from another (§11.2); (2) derive an expression in terms of $\boldsymbol{\nabla}$ or $\Gamma^{\alpha}{}_{\beta\gamma}$ for the "Riemann curvature tensor," which produces the geodesic deviation (§11.3); (3) see Riemann curvature at work, producing changes in vectors that are parallel transported around closed circuits (§11.4); (4) see Riemann curvature test whether spacetime is flat (§11.5); and (5) construct a special coordinate system, "Riemann normal coordinates," which is tied in a special way to the Riemann curvature tensor (§11.6).

§11.2. THE RELATIVE ACCELERATION OF NEIGHBORING GEODESICS

Geometry of a family of geodesics:

Selector parameter

Focus attention on a family of geodesics (Figure 11.1). Let one geodesic be distinguished from another by the value of a "selector parameter" n. The family includes not only geodesics $n = 0, 1, 2, \ldots$ but also geodesics for all intervening values of

2

Box 11.1 GEODESIC DEVIATION AND RIEMANN CURVATURE IN BRIEF

"Geodesic separation" $\boldsymbol{n}$ is displacement (tangent vector) from point on fiducial geodesic to point on nearby geodesic characterized by same value of affine parameter λ.

Geodesic separation changes with respect to λ (i.e., changes along the tangent vector $\boldsymbol{u} = d/d\lambda$) at a rate given by the *equation of geodesic deviation*

$$\nabla_{\boldsymbol{u}} \nabla_{\boldsymbol{u}} \boldsymbol{n} + \boldsymbol{Riemann}(\ldots, \boldsymbol{u}, \boldsymbol{n}, \boldsymbol{u}) = 0 \quad (1)$$

(second-order equation; see §§1.6 and 1.7; Figures 1.10, 1.11, 1.12).

In terms of components of the Riemann tensor the driving force ("tidal graviational force") is

$$\boldsymbol{Riemann}(\ldots, \boldsymbol{u}, \boldsymbol{n}, \boldsymbol{u}) = \boldsymbol{e}_\alpha R^\alpha{}_{\beta\gamma\delta} u^\beta n^\gamma u^\delta. \quad (2)$$

The components of the Riemann curvature tensor in a coordinate frame are given in terms of the connection coefficients by the formula

$$R^\alpha{}_{\beta\gamma\delta} = \frac{\partial \Gamma^\alpha{}_{\beta\delta}}{\partial x^\gamma} - \frac{\partial \Gamma^\alpha{}_{\beta\gamma}}{\partial x^\delta} + \Gamma^\alpha{}_{\mu\gamma}\Gamma^\mu{}_{\beta\delta} - \Gamma^\alpha{}_{\mu\delta}\Gamma^\mu{}_{\beta\gamma}. \quad (3)$$

This curvature tensor not only quantifies the concept of "tidal gravitational force," but also enters into Einstein's law, by which "matter tells spacetime how to curve." That law, to be studied in later chapters, takes the following operational-computational form in a given coordinate system:

(a) Write down trial formula for dynamic evolution of metric coefficients $g_{\mu\nu}$ with time.

(b) Calculate the connection coefficients from

$$\Gamma^\alpha{}_{\mu\nu} = g^{\alpha\beta}\Gamma_{\beta\mu\nu}; \quad (4)$$

$$\Gamma_{\beta\mu\nu} = \frac{1}{2}\left(\frac{\partial g_{\beta\nu}}{\partial x^\mu} + \frac{\partial g_{\beta\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\beta}\right) \quad (5)$$

(derived in Chapter 13).

(c) Calculate Riemann curvature tensor from equation (3).

(d) Calculate Einstein curvature tensor from

$$G_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu} - \frac{1}{2} g_{\mu\nu} g^{\sigma\tau} R^\alpha{}_{\sigma\alpha\tau} \quad (6)$$

(geometric significance in Chapter 15).

(e) Insert into Einstein's equations (Chapter 17):

$G_{\mu\nu} = 0$ (empty space),

$G_{\mu\nu} = 8\pi T_{\mu\nu}$ (when mass-energy is present).

(f) Test whether the trial formula for the dynamic evolution of the geometry was correct, and, if not, change it so it is.

Affine parameter

n. The typical point $\mathscr{P}$ on the typical geodesic will be a continuous, doubly differentiable function of the selector parameter n and the affine parameter λ; thus

$$\mathscr{P} = \mathscr{P}(\lambda, n). \quad (11.1)$$

Tangent vector

The tangent vector

$$\boldsymbol{u} = \frac{\partial \mathscr{P}}{\partial \lambda} \quad \text{(Cartan notation)}$$

or

$$\boldsymbol{u} = \frac{\partial}{\partial \lambda} \quad \text{(notation of this book)} \quad (11.2)$$

is constant along any given geodesic in this sense: the vector $\boldsymbol{u}$ at any point, trans-

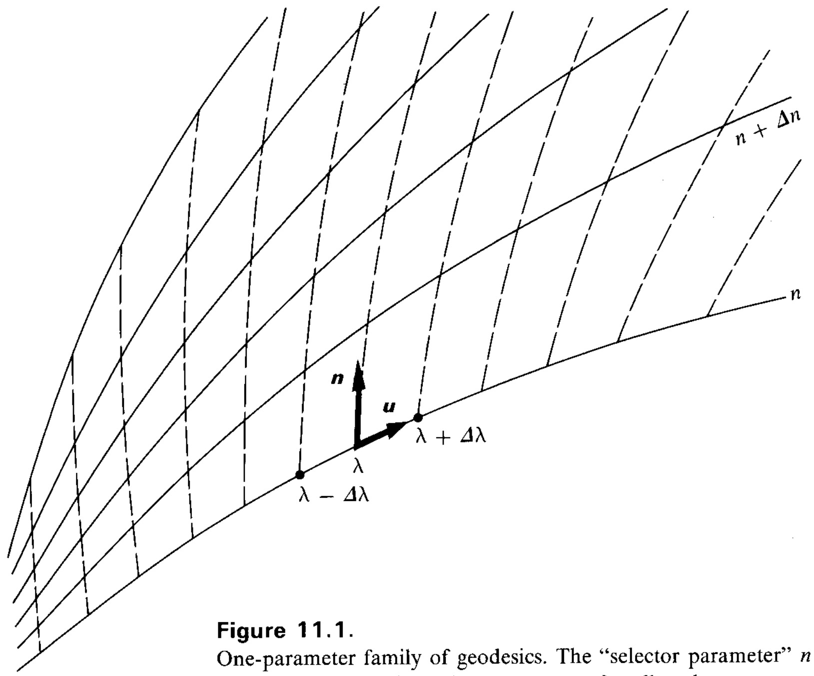


Figure 11.1.
One-parameter family of geodesics. The "selector parameter" n tells which geodesic. The affine parameter λ tells where on a given geodesic. The two tangent vectors indicated in the diagram are $\boldsymbol{u} = \partial/\partial\lambda$ (Cartan: $\partial\mathscr{P}/\partial\lambda$) and $\boldsymbol{n} = \partial/\partial n$ (Cartan: $\partial\mathscr{P}/\partial n$).

ported parallel to itself along the geodesic, arrives at a second point coincident in direction and length with the $\boldsymbol{u}$ already existing at that point.

Separation vector

The *"separation vector"*

$$\boldsymbol{n} = \frac{\partial\mathscr{P}}{\partial n} \qquad \text{(Cartan notation)}$$

or

$$\boldsymbol{n} = \frac{\partial}{\partial n} \qquad \text{(notation of this book)} \tag{11.3}$$

measures the separation between the geodesic n, regarded as the fiducial geodesic, and the typical nearby geodesic, $n + \Delta n$ (for small Δn), in the sense that

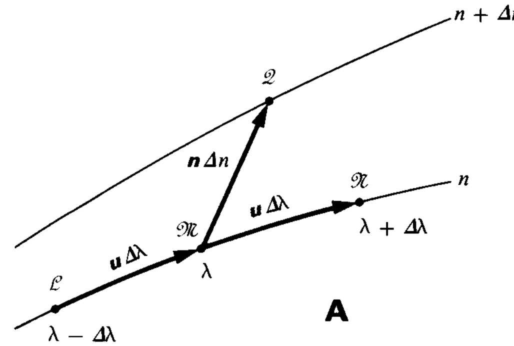
$$(\Delta n)\boldsymbol{n} = \begin{Bmatrix} \Delta n \dfrac{\partial\mathscr{P}}{\partial n} \\ \\ \Delta n \dfrac{\partial}{\partial n} \end{Bmatrix} \begin{matrix} \text{measures the} \\ \text{change in} \end{matrix} \begin{Bmatrix} \text{position} \\ \\ \text{any} \\ \text{function} \end{Bmatrix} \tag{11.4}$$

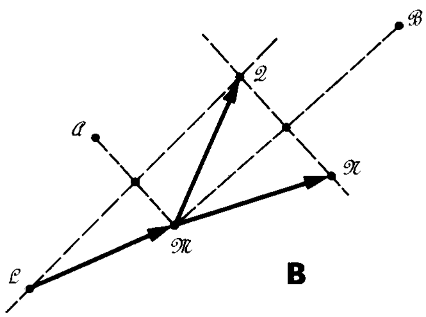
brought about by transfer of attention from the one geodesic to the other at a fixed value of the affine parameter λ. This vector is represented by the arrow $\mathscr{M}\mathscr{Q}$ in the first diagram in Box 11.2.

(continued on page 270)

Box 11.2 GEODESIC DEVIATION REPRESENTED AS AN ARROW

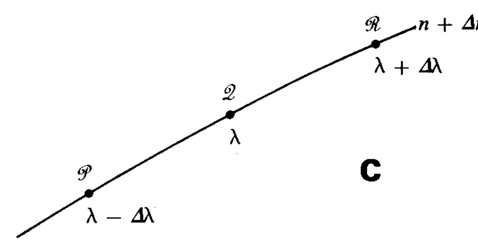
"Fiducial geodesic" n. Separation vector $\boldsymbol{n}\,\Delta n = \mathcal{M}\mathcal{Q}$ leads from point $\mathcal{M}$ on it, to point $\mathcal{Q}$ with same value of affine parameter λ (timelike quantity) on neighboring "test geodesic" $n + \Delta n$.

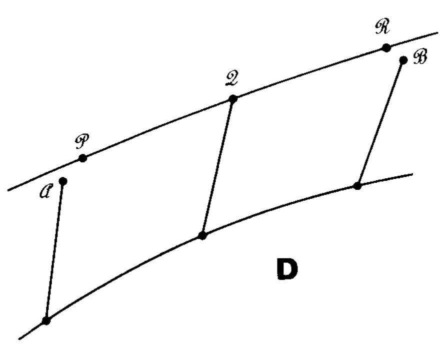




Parallel transport of $\mathcal{M}\mathcal{Q}$ by "Schild's ladder construction" (Box 10.2) to $\mathcal{N}\mathcal{B}$ and $\mathcal{L}\mathcal{A}$. If the test geodesic $n + \Delta n$ had kept a constant separation from the fiducial geodesic n, its tracer point would have arrived at $\mathcal{A}$ at the value $(\lambda - \Delta\lambda)$ of the affine parameter, and at $\mathcal{B}$ at $(\lambda + \Delta\lambda)$.

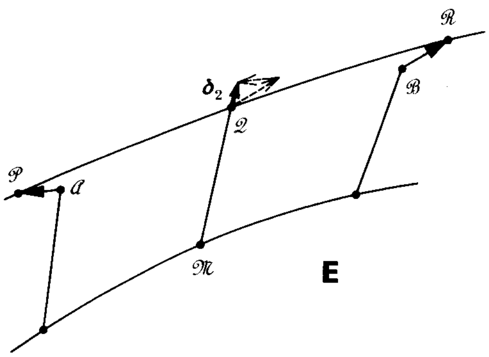
Actual location of tracer point of test geodesic at values of the timelike affine parameter $(\lambda - \Delta\lambda)$, λ, and $(\lambda + \Delta\lambda)$.



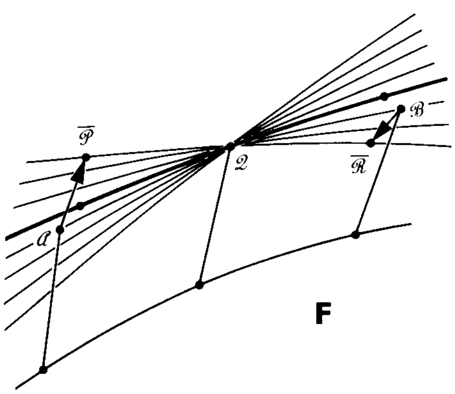


Confrontation between actual course of tracer point on test geodesic and "canonical course": course it would have had to take to keep constant separation from the tracer point moving along the fiducial geodesic.

Test geodesic same as before, except for uniform stretchout in scale of affine parameter. Any measure of departure of the actual course of geodesic from the canonical course ($\mathcal{A}\mathcal{Q}\mathcal{B}$), to be useful, should be independent of this stretchout. Hence, take as measure of geodesic deviation, not the vector $\mathcal{B}\mathcal{R}$ alone, nor the vector $\mathcal{A}\mathcal{P}$, but the stretch-independent combination $\boldsymbol{\delta}_2 = (\mathcal{B}\mathcal{R}) + (\mathcal{A}\mathcal{P})$. Here the sign of addition implies that the two vectors have been transported parallel to themselves, before addition, to a common point ($\mathcal{Q}$ in the diagram; $\mathcal{M}$ in the differential calculus limit $\Delta n \longrightarrow 0$, $\Delta\lambda \longrightarrow 0$).



E



F

Alternative courses that the test geodesic of **D** could have taken through $\mathcal{Q}$ (families of geodesics characterized by different degrees of divergence from the left or convergence towards the right). Tilt changes values of $\mathcal{A}\mathcal{P}$ (to $\mathcal{A}\overline{\mathcal{P}}$) and $\mathcal{B}\mathcal{R}$ (to $\mathcal{B}\overline{\mathcal{R}}$) individually, but not value of the sum $\boldsymbol{\delta}_2 = (\mathcal{B}\mathcal{R}) + (\mathcal{A}\mathcal{P})$ ("lever principle").

Note that arrow $\mathcal{B}\mathcal{R}$ is of first order in $\Delta\lambda$ and of first order in Δn; similarly for $\mathcal{A}\mathcal{P}$; hence the combination $\boldsymbol{\delta}_2$ is of second order in $\Delta\lambda$ and first order in Δn. Conclude that *the arrow* $\boldsymbol{\delta}_2/(\Delta\lambda)^2(\Delta n)$ *is the desired measure of geodesic deviation* in the sense that:

size of mesh (ultimately to go to zero) cancels out;
parameterization of test geodesic cancels out;
slope of test geodesic cancels out.

Give this arrow the name *"relative-acceleration vector";* and by examining it more closely (Box 11.3), discover the formula

$$\boldsymbol{\delta}_2/(\Delta\lambda)^2(\Delta n) = \boldsymbol{\nabla}_{\boldsymbol{u}}\,\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{n}$$

for it.

Box 11.2 illustrates what it means to speak of geodesic deviation. One transports the separation $\boldsymbol{n}\,\Delta n = \mathcal{M}\mathcal{Q}$ parallel to itself along the fiducial geodesic. The tip of this vector traces out the canonical course that the nearby tracer point would have to pursue if it were to maintain constant separation from the fiducial tracer point. The actual course of the test geodesic deviates from this "canonical" course. The deviation, a vector ($\mathcal{A}\mathcal{P}$ of Box 11.2), changes with the affine parameter ($\mathcal{A}\mathcal{P}$ at $\mathcal{A}$, 0 at $\mathcal{Q}$, $\mathcal{B}\mathcal{R}$ at $\mathcal{B}$). The first derivative of this vector with respect to the affine parameter is sensitive to the scale of parameterization along the test geodesic, and to its slope (Box 11.2, **F**). Not so the second derivative. It depends only on the tangent vector $\boldsymbol{u}$ of the fiducial geodesic, and on the separation vector $\boldsymbol{n}\,\Delta n$. Divide this second derivative of the deviation by Δn and give it a name: the *"relative-acceleration vector"*. Discover (Box 11.3) a simple formula for it

Relative-acceleration vector

$$(\text{relative-acceleration vector}) = \nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n}. \tag{11.5}$$

§11.3. TIDAL GRAVITATIONAL FORCES AND RIEMANN CURVATURE TENSOR

With "relative acceleration" now defined, turn to the "tidal gravitational force" (i.e., "spacetime curvature") that produces it. Use a Newtonian analysis of tidal forces

Box 11.3 GEODESIC DEVIATION: ARROW CORRELATED WITH SECOND COVARIANT DERIVATIVE

The arrow $\boldsymbol{\delta}_2$ in Box 11.2 measures, not the rate of change of the separation of the test geodesic $n + \Delta n$ from the "canonical course" $\mathcal{A}\mathcal{Q}\mathcal{B}$ as baseline, but the second derivative:

$$\left(\text{first derivative at } \lambda + \frac{1}{2}\Delta\lambda\right) = \nabla_{\boldsymbol{u}}\boldsymbol{n} = \frac{\mathcal{N}\mathcal{R} - \mathcal{N}\mathcal{B}}{\Delta\lambda\,\Delta n} = \frac{\mathcal{B}\mathcal{R}}{\Delta\lambda\,\Delta n};$$

$$\left(\text{first derivative at } \lambda - \frac{1}{2}\Delta\lambda\right) = \nabla_{\boldsymbol{u}}\boldsymbol{n} = \frac{\mathcal{L}\mathcal{A} - \mathcal{L}\mathcal{P}}{\Delta\lambda\,\Delta n} = \frac{-\mathcal{A}\mathcal{P}}{\Delta\lambda\,\Delta n}.$$

Transpose to common location λ, take difference, and divide it by $\Delta\lambda$ to obtain the second covariant derivative with respect to the vector $\boldsymbol{u}$; thus

$$\begin{aligned}\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n} &= \frac{(\nabla_{\boldsymbol{u}}\boldsymbol{n})_{\lambda+\frac{1}{2}\Delta\lambda} - (\nabla_{\boldsymbol{u}}\boldsymbol{n})_{\lambda-\frac{1}{2}\Delta\lambda}}{\Delta\lambda}\\ &= \frac{(\mathcal{B}\mathcal{R} + \mathcal{A}\mathcal{P})_{\substack{\text{vectors transported to}\\ \text{common location}}}}{(\Delta\lambda)^2\,\Delta n} = \frac{\boldsymbol{\delta}_2}{(\Delta\lambda)^2\,\Delta n}\\ &= \text{"relative acceleration vector" for neighboring geodesics.}\end{aligned}$$

(left half of Box 11.4) to motivate the geometric analysis (right half of same box). Thereby arrive at the remarkable equation

Tide-producing gravitational forces expressed in terms of a commutator

$$\underbrace{\nabla_u \nabla_u \boldsymbol{n}}_{\text{“relative acceleration”}} \quad + \quad \underbrace{[\nabla_n, \nabla_u]\boldsymbol{u}}_{\text{“tide-producing gravitational forces”}} = 0. \tag{11.6}$$

This equation is remarkable, because at first sight it seems crazy. The term $[\nabla_n, \nabla_u]\boldsymbol{u}$ involves second derivatives of $\boldsymbol{u}$, and a first derivative of ∇_n:

$$[\nabla_n, \nabla_u]\boldsymbol{u} \equiv \nabla_n \nabla_u \boldsymbol{u} - \nabla_u \nabla_n \boldsymbol{u}. \tag{11.7}$$

It thus must depend on how $\boldsymbol{u}$ and $\boldsymbol{n}$ vary from point to point. But the relative acceleration it produces, $\nabla_u \nabla_u \boldsymbol{n}$, is known to depend only on the values of $\boldsymbol{u}$ and $\boldsymbol{n}$ at the fiducial point, not on how $\boldsymbol{u}$ and $\boldsymbol{n}$ vary (see Box 11.2, **F**). How is this possible?

Somehow all derivatives must drop out of the tidal-force quantity $[\nabla_n, \nabla_u]\boldsymbol{u}$. One must be able to regard $[\nabla_{\dots}, \nabla_{\dots}]\dots$ as a purely local, algebraic machine with three slots, whose output is a vector. If it is purely local and not differential, then it is even linear (as one sees from the additivity properties of ∇), so it must be a tensor. Give this tensor the name ***Riemann***, and give it a fourth slot for inputting a 1-form:

$$\textbf{\textit{Riemann}}\,(\dots, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) \equiv [\nabla_A, \nabla_B]\boldsymbol{C};$$

$$\textbf{\textit{Riemann}}\,(\boldsymbol{\sigma}, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) \equiv \langle \boldsymbol{\sigma}, [\nabla_A, \nabla_B]\boldsymbol{C}\rangle.$$

This is only a tentative definition of ***Riemann***. Before accepting it, one should verify that it is, indeed, a tensor. Does it *really* depend on only the values of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ at the point of evaluation, and not on how they are changing there? The answer (derived in Box 11.5) is "almost." It fails the test, but with a slight modification it will pass. The modification is to replace the commutator $[\nabla_A, \nabla_B]$ by the *"curvature operator"*

Curvature operator defined

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{B}) \equiv [\nabla_A, \nabla_B] - \nabla_{[A,B]}, \tag{11.8}$$

where $\nabla_{[A,B]}$ is the derivative along the vector $[\boldsymbol{A}, \boldsymbol{B}]$ (commutator of $\boldsymbol{A}$ and $\boldsymbol{B}$). ($\mathcal{R}(\boldsymbol{A}, \boldsymbol{B}) \equiv [\nabla_A, \nabla_B]$ for the fields $\boldsymbol{A} = \boldsymbol{n}$ and $\boldsymbol{B} = \boldsymbol{u}$ of the geodesic-deviation problem, because $[\boldsymbol{n}, \boldsymbol{u}] = 0$.) Then the modified and acceptable *definition of the Riemann curvature tensor is*

Riemann curvature tensor defined

$$\begin{aligned} \textbf{\textit{Riemann}}\,(\dots, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) &\equiv \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}; \\ \textbf{\textit{Riemann}}\,(\boldsymbol{\sigma}, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) &\equiv \langle \boldsymbol{\sigma}, \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}\rangle. \end{aligned} \tag{11.9}$$

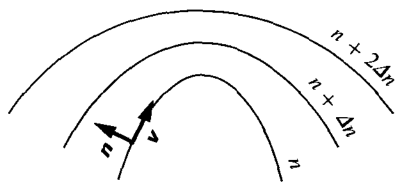
To define ***Riemann*** thus, and to verify its tensorial character (exercise 11.2), does not by any means teach one what curvature is all about. To understand curvature, one must scrutinize ***Riemann*** from all viewpoints. That is the task of the rest of this chapter.

(*continued on page 275*)

Box 11.4 RELATIVE ACCELERATION OF TEST PARTICLES—GEOMETRIC ANALYSIS PATTERNED ON NEWTONIAN ANALYSIS

Newtonian Analysis

1. Consider a family of test-particle trajectories $x^j(t, n)$ in ordinary, three-dimensional space: "t" is time measured by particle's clock, or any clock; "n" is "selector parameter."



2. Equation of motion for each trajectory:

$$\left(\frac{\partial^2 x^j}{\partial t^2}\right)_n + \frac{\partial \Phi}{\partial x^j} = 0,$$

where Φ is Newtonian potential.

3. Take difference between equations of motion for neighboring trajectories, n and $n + \Delta n$, and take limit as $\Delta n \longrightarrow 0$—i.e., take derivative

$$\left(\frac{\partial}{\partial n}\right)_t \left[\left(\frac{\partial^2 x^j}{\partial t^2}\right)_n + \frac{\partial \Phi}{\partial x^j}\right] = 0.$$

4. When $\partial/\partial n$ acts on second term, rewrite it as

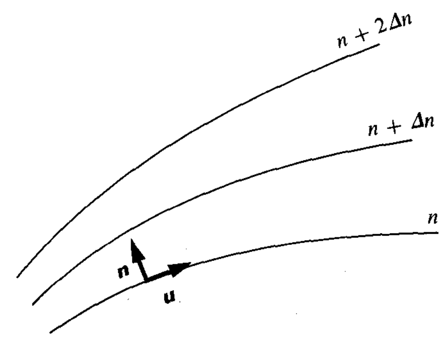
$$\left(\frac{\partial}{\partial n}\right)_t = \left(\frac{\partial x^k}{\partial n}\right)_t \frac{\partial}{\partial x^k} = n^k \frac{\partial}{\partial x^k};$$

Thereby obtain

$$\left(\frac{\partial}{\partial n}\right)_t \left(\frac{\partial}{\partial t}\right)_n \left(\frac{\partial x^j}{\partial t}\right)_n + \frac{\partial^2 \Phi}{\partial x^j \partial x^k} n^k = 0.$$

Geometric Analysis

1. Consider a family of test-particle trajectories (geodesics), $\mathscr{P}(\lambda, n)$, in spacetime: "λ" is affine-parameter, i.e., time measured by particle's clock; "n" is "selector parameter."



2. Geodesic equation for each trajectory:

$$\nabla_{\boldsymbol{u}} \boldsymbol{u} = 0.$$

[Looks like first-order equation; is actually second-order because the "$\boldsymbol{u}$" being differentiated is itself a derivative, $\boldsymbol{u} = (\partial \mathscr{P}/\partial \lambda)_n$.]

3. Take difference between geodesic equations for neighboring geodesics n and $n + \Delta n$, and take limit as $\Delta n \longrightarrow 0$—i.e., take covariant derivative

$$\nabla_{\boldsymbol{n}} [\nabla_{\boldsymbol{u}} \boldsymbol{u}] = 0.$$

4. There is no second term, so leave equation in form

$$\nabla_{\boldsymbol{n}} [\nabla_{\boldsymbol{u}} \boldsymbol{u}] = 0.$$

5. To obtain equation for relative acceleration, move $(\partial/\partial n)_t$ through both of the $(\partial/\partial t)_n$ terms (permissible because partial derivatives commute!):

$$\left(\frac{\partial}{\partial t}\right)_n\left(\frac{\partial}{\partial t}\right)_n\left(\frac{\partial x^j}{\partial n}\right)_t + \frac{\partial^2\Phi}{\partial x^j\,\partial x^k}n^k = 0.$$

This is equivalent to

$$\boxed{\underbrace{\left(\frac{\partial^2 n^j}{\partial t^2}\right)}_{\text{"relative acceleration"}} + \underbrace{\frac{\partial^2\Phi}{\partial x^j\,\partial x^k}n^k}_{\text{"tide-producing gravitational forces"}} = 0.}$$

5. To obtain equation for relative acceleration, $\nabla_u\nabla_u n$, move ∇_n through ∇_u and through the $\partial/\partial\lambda$ of $u = \partial\mathscr{P}/\partial\lambda$:

a. *First step:* In $\nabla_n\nabla_u u = 0$, move ∇_n through ∇_u. The result:

$$(\nabla_u\nabla_n + \underbrace{[\nabla_n, \nabla_u]}_{\text{commutator; must be included as protection against possibility that } \nabla_u\nabla_n \neq \nabla_n\nabla_u.})u = 0.$$

b. *Second step:* Move ∇_n through $\partial/\partial\lambda$ of $u = \partial\mathscr{P}/\partial\lambda$; i.e., write

$$\nabla_n\frac{\partial\mathscr{P}}{\partial\lambda} \underset{\text{def. of } u}{=} \nabla_n u \underset{\text{Why?}}{=} \nabla_u n \underset{\text{def. of } n}{=} \nabla_u\frac{\partial\mathscr{P}}{\partial n}$$

Why? Because symmetry of covariant derivative says $\nabla_n u - \nabla_u n = [n, u]$

$$= \left[\frac{\partial}{\partial n}, \frac{\partial}{\partial\lambda}\right] = \frac{\partial^2}{\partial n\,\partial\lambda} - \frac{\partial^2}{\partial\lambda\,\partial n} = 0.$$

c. Result:

$$\boxed{\underbrace{\nabla_u\nabla_u n}_{\text{"relative acceleration"}} + \underbrace{[\nabla_n, \nabla_u]u}_{\text{"tide-producing gravitational forces"; i.e., "spacetime curvature"}} = 0}$$

Box 11.5 RIEMANN CURVATURE TENSOR

A. Definition of *Riemann* Motivated by Tidal Gravitational Forces:

1. Tidal forces (spacetime curvature) produce relative acceleration of test particles (geodesics) given by

$$\nabla_u\nabla_u n + [\nabla_n, \nabla_u]u = 0. \tag{1}$$

Box 11.5 (*continued*)

2. This motivates the definition

$$\textbf{\textit{Riemann}}\,(\underset{\uparrow}{\ldots}, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) = [\boldsymbol{\nabla}_{\boldsymbol{A}}, \boldsymbol{\nabla}_{\boldsymbol{B}}]\boldsymbol{C}. \tag{2}$$

[empty slot for inserting a one-form]

B. Failure of this Definition

1. Definition acceptable only if ***Riemann*** $(\ldots, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B})$ is a linear machine, independent of how $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ vary from point to point.
2. Check, in part: change variations of $\boldsymbol{C}$, but not $\boldsymbol{C}$ itself, at event $\mathscr{P}_0$:

$$\boldsymbol{C}_{\mathrm{NEW}}(\mathscr{P}) = \underset{\uparrow}{f}(\mathscr{P})\boldsymbol{C}_{\mathrm{OLD}}(\mathscr{P}).$$

[arbitrary function except $f(\mathscr{P}_0) = 1$]

3. Does this change $[\boldsymbol{\nabla}_{\boldsymbol{A}}, \boldsymbol{\nabla}_{\boldsymbol{B}}]\boldsymbol{C}$? *Yes!* Exercise 11.1 shows

$$\{[\boldsymbol{\nabla}_{\boldsymbol{A}}, \boldsymbol{\nabla}_{\boldsymbol{B}}]\boldsymbol{C}_{\mathrm{NEW}}\}_{\mathrm{at}\,\mathscr{P}_0} - \{[\boldsymbol{\nabla}_{\boldsymbol{A}}, \boldsymbol{\nabla}_{\boldsymbol{B}}]\boldsymbol{C}_{\mathrm{OLD}}\}_{\mathscr{P}_0} = \boldsymbol{C}_{\mathrm{OLD}}\,\boldsymbol{\nabla}_{[\boldsymbol{A},\boldsymbol{B}]}f.$$

C. Modified Definition of ***Riemann***:

1. The term causing trouble, $\boldsymbol{C}_{\mathrm{OLD}}\,\boldsymbol{\nabla}_{[\boldsymbol{A},\boldsymbol{B}]}f$, can be disposed of by subtracting a "correction term" resembling it from ***Riemann***—i.e., by redefining

$$\textbf{\textit{Riemann}}\,(\ldots, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) \equiv \mathscr{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}, \tag{3}$$

$$\mathscr{R}(\boldsymbol{A}, \boldsymbol{B}) \equiv [\boldsymbol{\nabla}_{\boldsymbol{A}}, \boldsymbol{\nabla}_{\boldsymbol{B}}] - \boldsymbol{\nabla}_{[\boldsymbol{A},\boldsymbol{B}]} \tag{4}$$

2. The above calculation then gives a result independent of the "modifying function" f:

$$\{\mathscr{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}_{\mathrm{NEW}}\}_{\mathrm{at}\,\mathscr{P}_0} = \{\mathscr{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}_{\mathrm{OLD}}\}_{\mathrm{at}\,\mathscr{P}_0}.$$

D. Is Modified Definition Compatible with Equation for Tidal Gravitational Forces?

1. One would like to write $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{n} + \textbf{\textit{Riemann}}\,(\ldots, \boldsymbol{u}, \boldsymbol{n}, \boldsymbol{u}) = 0$.
2. This works just as well for modified definition of ***Riemann*** as for original definition, because

$$\mathcal{R}(\boldsymbol{n}, \boldsymbol{u}) = [\boldsymbol{\nabla}_{\boldsymbol{n}}, \boldsymbol{\nabla}_{\boldsymbol{u}}] - \underbrace{\boldsymbol{\nabla}_{[\boldsymbol{n},\boldsymbol{u}]}}_{\begin{bmatrix}= 0 \text{ because } \boldsymbol{n} = (\partial/\partial n)_\lambda \text{ and} \\ \boldsymbol{u} = (\partial/\partial\lambda)_n \text{ commute}\end{bmatrix}} = [\boldsymbol{\nabla}_{\boldsymbol{n}}, \boldsymbol{\nabla}_{\boldsymbol{u}}].$$

Geodesic deviation and tidal forces cannot tell the difference between $\mathcal{R}(\boldsymbol{n}, \boldsymbol{u})$ and $[\boldsymbol{\nabla}_{\boldsymbol{n}}, \boldsymbol{\nabla}_{\boldsymbol{u}}]$, nor consequently between old and new definitions of ***Riemann***.

E. Is Modified Definition Acceptable?

I.e., is ***Riemann*** $(\ldots, \boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) \equiv \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}$ a linear machine with output independent of how $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ vary near point of evaluation? YES! (See exercise 11.2.)

Take stock, first, of what one knows already about the Riemann curvature tensor. (1) ***Riemann*** *is* a tensor; despite the appearance of $\boldsymbol{\nabla}$ in its definition (11.9), no derivatives actually act on the input vectors $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$. (2) ***Riemann*** is a $\binom{1}{3}$ tensor; its first slot accepts a 1-form; the others, vectors. (3) ***Riemann*** is determined entirely by $\boldsymbol{\nabla}$, or equivalently by the geodesics of spacetime, or equivalently by spacetime's parallel transport law; nothing but $\boldsymbol{\nabla}$ and the input vectors and 1-form are required to fix ***Riemann***'s output. (4) ***Riemann*** produces the tidal gravitational forces that pry geodesics (test-particle trajectories) apart or push them together; i.e., it characterizes the "curvature of spacetime":

Tide-producing gravitational forces expressed in terms of ***Riemann***

$$\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{n} + \boldsymbol{Riemann}\,(\ldots, \boldsymbol{u}, \boldsymbol{n}, \boldsymbol{u}) = 0. \tag{11.10}$$

(This *"equation of geodesic deviation"* follows from equations 11.6, 11.8, and 11.9, and the relation $[\boldsymbol{n}, \boldsymbol{u}] = 0$.)

All these facets of ***Riemann*** are *pictorial* (e.g., geodesic deviation; see Boxes 11.2 and 11.3) or *abstract* (e.g., equations 11.8 and 11.9 for ***Riemann*** in terms of $\boldsymbol{\nabla}$). ***Riemann***'s *component* facet,

Components of ***Riemann*** expressed in terms of connection coefficients

$$R^\alpha{}_{\beta\gamma\delta} \equiv \boldsymbol{Riemann}\,(\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta, \boldsymbol{e}_\gamma, \boldsymbol{e}_\delta) \equiv \langle \boldsymbol{\omega}^\alpha, \mathcal{R}(\boldsymbol{e}_\gamma, \boldsymbol{e}_\delta)\, \boldsymbol{e}_\beta \rangle, \tag{11.11}$$

is related to the component facet of $\boldsymbol{\nabla}$ by the following equation, valid in any coordinate basis $\{\boldsymbol{e}_\alpha\} = \{\partial/\partial x^\alpha\}$:

$$R^\alpha{}_{\beta\gamma\delta} = \frac{\partial \Gamma^\alpha{}_{\beta\delta}}{\partial x^\gamma} - \frac{\partial \Gamma^\alpha{}_{\beta\gamma}}{\partial x^\delta} + \Gamma^\alpha{}_{\mu\gamma}\Gamma^\mu{}_{\beta\delta} - \Gamma^\alpha{}_{\mu\delta}\Gamma^\mu{}_{\beta\gamma}. \tag{11.12}$$

(See exercise 11.3 for derivation, and exercise 11.4 for the extension to noncoordinate bases.) These components of ***Riemann***, with no sign of any derivative operator anywhere, may leave one with a better feeling in one's stomach than the definition (11.8) with its nondifferentiating derivatives!

EXERCISES

Exercise 11.1. $[\nabla_A, \nabla_B]C$ DEPENDS ON DERIVATIVES OF C

(Based on Box 11.5.) Let $\boldsymbol{C}_{\rm NEW}$ and $\boldsymbol{C}_{\rm OLD}$ be vector fields related by

$$\boldsymbol{C}_{\rm NEW}(\mathcal{P}) = \underbrace{f(\mathcal{P})}_{\text{[arbitrary function, except } f(\mathcal{P}_0) = 1]}\boldsymbol{C}_{\rm OLD}(\mathcal{P}).$$

Show that

$$\{[\nabla_{\boldsymbol{A}}, \nabla_{\boldsymbol{B}}]\boldsymbol{C}_{\rm NEW}\}_{{\rm at}\,\mathcal{P}_0} - \{[\nabla_{\boldsymbol{A}}, \nabla_{\boldsymbol{B}}]\boldsymbol{C}_{\rm OLD}\}_{{\rm at}\,\mathcal{P}_0} = \boldsymbol{C}_{\rm OLD}\,\nabla_{[\boldsymbol{A},\boldsymbol{B}]}f.$$

Exercise 11.2. PROOF THAT *Riemann* IS A TENSOR

Show from its definition (11.8, 11.9) that ***Riemann*** is a tensor. Hint: Use the following procedure.

(a) If $f(\mathcal{P})$ is an arbitrary function, show that

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})f\boldsymbol{C} = f\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}.$$

(b) Similarly show that

$$\mathcal{R}(f\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C} = f\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C} \quad \text{and} \quad \mathcal{R}(\boldsymbol{A}, f\boldsymbol{B})\boldsymbol{C} = f\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}.$$

(c) Show that $\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}$ is linear; i.e.,

$$\mathcal{R}(\boldsymbol{A} + \boldsymbol{a}, \boldsymbol{B})\boldsymbol{C} = \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C} + \mathcal{R}(\boldsymbol{a}, \boldsymbol{B})\boldsymbol{C};$$
$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{B} + \boldsymbol{b})\boldsymbol{C} = \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C} + \mathcal{R}(\boldsymbol{A}, \boldsymbol{b})\boldsymbol{C};$$
$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})(\boldsymbol{C} + \boldsymbol{c}) = \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C} + \mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{c}.$$

(d) Now use the above properties to prove the most crucial feature of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}$: Modify the variations (gradients) of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ in an arbitrary manner, but leave $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ unchanged at $\mathcal{P}_0$:

$$\left.\begin{array}{l} \boldsymbol{A} \longrightarrow \boldsymbol{A} + a^\alpha \boldsymbol{e}_\alpha \\ \boldsymbol{B} \longrightarrow \boldsymbol{B} + b^\alpha \boldsymbol{e}_\alpha \\ \boldsymbol{C} \longrightarrow \boldsymbol{C} + c^\alpha \boldsymbol{e}_\alpha \end{array}\right\} \quad \begin{array}{l} a^\alpha(\mathcal{P}),\ b^\alpha(\mathcal{P}),\ c^\alpha(\mathcal{P}) \text{ arbitrary except} \\ \text{they all vanish at } \mathcal{P} = \mathcal{P}_0. \end{array}$$

Show that this modification leaves $\mathcal{R}(\boldsymbol{A}, \boldsymbol{B})\boldsymbol{C}$ unchanged at $\mathcal{P}_0$.
(e) From these facts, conclude that ***Riemann*** is a tensor.

Exercise 11.3. COMPONENTS OF *Riemann* IN COORDINATE BASIS

Derive equation (11.12) for the components of the Riemann tensor in a coordinate basis. [*Solution:*

$$\begin{aligned} R^\alpha{}_{\beta\gamma\delta} &= \textit{\textbf{Riemann}}\,(\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta, \boldsymbol{e}_\gamma, \boldsymbol{e}_\delta) && \begin{bmatrix}\text{standard way to} \\ \text{calculate components}\end{bmatrix} \\ &= \langle \boldsymbol{\omega}^\alpha, \mathcal{R}(\boldsymbol{e}_\gamma, \boldsymbol{e}_\delta)\boldsymbol{e}_\beta \rangle && \text{[by definition (11.9)]} \\ &= \langle \boldsymbol{\omega}^\alpha, (\nabla_\gamma \nabla_\delta - \nabla_\delta \nabla_\gamma)\boldsymbol{e}_\beta \rangle && \begin{bmatrix}\text{by definition (11.8) plus} \\ [\boldsymbol{e}_\gamma, \boldsymbol{e}_\delta] = 0 \text{ in coord. basis}\end{bmatrix} \\ &= \langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\mu \Gamma^\mu{}_{\beta\delta,\gamma} + (\boldsymbol{e}_\nu \Gamma^\nu{}_{\mu\gamma})\Gamma^\mu{}_{\beta\delta} - \boldsymbol{e}_\mu \Gamma^\mu{}_{\beta\gamma,\delta} - (\boldsymbol{e}_\nu \Gamma^\nu{}_{\mu\delta})\Gamma^\mu{}_{\beta\gamma} \rangle \\ &= (\Gamma^\mu{}_{\beta\delta,\gamma} - \Gamma^\mu{}_{\beta\gamma,\delta})\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\mu \rangle + (\Gamma^\nu{}_{\mu\gamma}\Gamma^\mu{}_{\beta\delta} - \Gamma^\nu{}_{\mu\delta}\Gamma^\mu{}_{\beta\gamma})\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\nu \rangle, \end{aligned}$$

which reduces (upon using $\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\mu \rangle = \delta^\alpha{}_\mu$) to (11.12).]

Exercise 11.4. COMPONENTS OF *RIEMANN* IN NONCOORDINATE BASIS

In a noncoordinate basis with commutation coefficients $c_{\alpha\beta}{}^{\gamma}$ defined by equation (9.22), derive the following equation for the components of ***Riemann***:

$$R^{\alpha}{}_{\beta\gamma\delta} = \Gamma^{\alpha}{}_{\beta\delta,\gamma} - \Gamma^{\alpha}{}_{\beta\gamma,\delta} + \Gamma^{\alpha}{}_{\mu\gamma}\Gamma^{\mu}{}_{\beta\delta} - \Gamma^{\alpha}{}_{\mu\delta}\Gamma^{\mu}{}_{\beta\gamma} - \Gamma^{\alpha}{}_{\beta\mu}c_{\gamma\delta}{}^{\mu}. \qquad (11.13)$$

§11.4. PARALLEL TRANSPORT AROUND A CLOSED CURVE

What are the effects of spacetime curvature, and how can one quantify them? One effect is geodesic deviation (relative acceleration of test bodies), quantified by equation (11.10). Another effect, almost as important, is the change in a vector caused by parallel transport around a closed curve. This effect shows up most clearly in the same problem, geodesic deviation, that motivated curvature in the first place. The relative acceleration vector $\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n}$ is also the change $\delta\boldsymbol{u}$ in the vector $\boldsymbol{u}$ caused by parallel transport around the curve whose legs are the vectors $\boldsymbol{n}$ and $\boldsymbol{u}$:

Change in a vector due to parallel transport around a closed curve:

Related to geodesic deviation

$$\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n} = \delta\boldsymbol{u}.$$

(See Box 11.6 for proof.) Hence, in this special case one can write

$$\delta\boldsymbol{u} + \textbf{Riemann}\,(\ldots,\boldsymbol{u},\boldsymbol{n},\boldsymbol{u}) = 0.$$

The expected generalization is obvious: pick a closed quadrilateral with legs $\boldsymbol{u}\,\Delta a$ and $\boldsymbol{v}\,\Delta b$ (Figure 11.2; Δa and Δb are small parameters, to go to zero at end of discussion). Parallel transport the vector $\boldsymbol{A}$ around this quadrilateral. The resultant change in $\boldsymbol{A}$ should satisfy the equation

$$\delta\boldsymbol{A} + \textbf{Riemann}\,(\ldots,\boldsymbol{A},\boldsymbol{u}\,\Delta a,\boldsymbol{v}\,\Delta b) = 0; \qquad (11.14)$$

Equation for change

or, equivalently,

$$\delta\boldsymbol{A} + \Delta a\,\Delta b\,\mathscr{R}(\boldsymbol{u},\boldsymbol{v})\boldsymbol{A} = 0; \qquad (11.14')$$

or, more precisely,

$$\lim_{\substack{\Delta a\to 0\\ \Delta b\to 0}}\left(\frac{\delta\boldsymbol{A}}{\Delta a\,\Delta b}\right) + \textbf{Riemann}\,(\ldots,\boldsymbol{A},\boldsymbol{u},\boldsymbol{v}) = 0. \qquad (11.14'')$$

The proof is enlightening, for it reveals the geometric origin of the correction term $\nabla_{[\boldsymbol{u},\boldsymbol{v}]}$ in the curvature operator.

The circuit of transport (Figure 11.2) is to be made from two arbitrary vector fields $\boldsymbol{u}\,\Delta a$ and $\boldsymbol{v}\,\Delta b$. However, a circuit made only of these fields has a gap in it, for a simple reason. The magnitude of $\boldsymbol{u}$ varies the wrong way from place to place. The displacement $\boldsymbol{u}\,\Delta a$ that reaches across at the bottom of the quadrilateral from

Derivation of equation for change

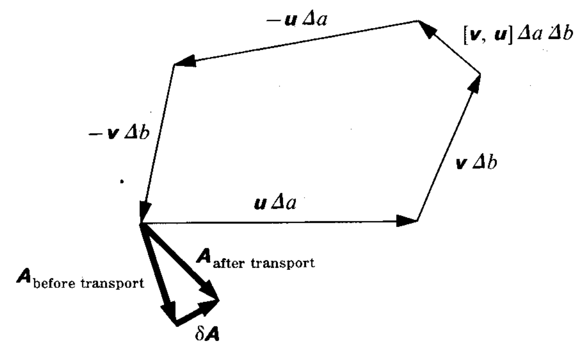


Figure 11.2.
The change $\delta\boldsymbol{A}$ in a vector $\boldsymbol{A}$ as a result of parallel transport around a closed curve. The edges of the curve are the vector fields $\boldsymbol{u}\,\Delta a$ and $\boldsymbol{v}\,\Delta b$, plus the "closer of the quadrilateral" $[\boldsymbol{v}\,\Delta b, \boldsymbol{u}\,\Delta a] = [\boldsymbol{v}, \boldsymbol{u}]\,\Delta a\,\Delta b$ (see Box 9.2).

one line of $\boldsymbol{v}$'s to another cannot make the connection at the top of the quadrilateral. Similarly the $\boldsymbol{v}$'s vary the wrong way from place to place to connect the $\boldsymbol{u}$'s. To close the gap and complete the circuit, insert the "closer of quadrilaterals" $[\boldsymbol{v}\,\Delta b, \boldsymbol{u}\,\Delta a] = [\boldsymbol{v}, \boldsymbol{u}]\,\Delta a\,\Delta b$. (See Box 9.2 for why this vector closes the gap.)

With the route now specified, the vector $\boldsymbol{A}$ is to be transported around it. One way to do this, "geometrical construction" by the method of Schild's ladder applied over and over, is the foundation for planning a possible experiment. For planning an abstract and coordinate-free calculation (the present line of action), introduce a "fiducial field," only to take it away at the end of the calculation. *Plan:* Conceive of $\boldsymbol{A}$, not as a localized vector defined solely at the start of the trip, but as a vector field (defined throughout the trip). *Purpose:* To provide a standard of reference (comparison of $\boldsymbol{A}$ transported from the origin with $\boldsymbol{A}$ at the place in question). *Principle:* The standard of reference will cancel out in the end. *Procedure:*

$-\delta\boldsymbol{A} = -$ (Net change made in taking the vector $\boldsymbol{A}$, originally localized at the start of the circuit, and transporting it parallel to itself ("mobile $\boldsymbol{A}$") around the closed circuit. This quantity cannot be evaluated until completion of circuit because there is no preexisting standard of reference along the way.)

$= +$ (A quantity subject to analysis for each leg of circuit individually. This new quantity is defined by introducing throughout the whole region a vector field $\boldsymbol{A}^{(\text{field})}$, smoothly varying, and in agreement at starting point with the original localized $\boldsymbol{A}$, but otherwise arbitrary. This new quantity is then given by $\boldsymbol{A}^{(\text{field})}$ at starting point (same as $\boldsymbol{A}^{(\text{localized})}$ at starting point) minus $\boldsymbol{A}^{(\text{mobile})}$ at finish point (after transit).)

$= \sum_{\text{legs of circuit}}$ (Change in $\boldsymbol{A}^{(\text{field})}$ relative to $\boldsymbol{A}^{(\text{mobile})}$ in the course of transport along specified leg. Value for any one leg depends on the arbitrary choice of $\boldsymbol{A}^{(\text{field})}$, but this arbitrariness cancels out in end because of closure of circuit.)

$$
= \left(\begin{array}{l}
\text{Change in } \boldsymbol{A}^{(\text{field})} \text{ relative to the parallel-transported } \boldsymbol{A}^{(\text{mobile})} \text{ as standard} \\
\text{of reference, made up of contributions along following legs of Figure} \\
\text{11.2:} \\
\quad \boldsymbol{v}\,\Delta b, \text{ giving } \quad \nabla_{\boldsymbol{v}}\boldsymbol{A}^{(\text{field})}\,\Delta b \text{ (on line displaced } \boldsymbol{u}\,\Delta a \text{ from start)} \\
-\boldsymbol{v}\,\Delta b, \text{ giving } -\nabla_{\boldsymbol{v}}\boldsymbol{A}^{(\text{field})}\,\Delta b \text{ (on line through starting point)} \\
-\boldsymbol{u}\,\Delta a, \text{ giving } -\nabla_{\boldsymbol{u}}\boldsymbol{A}^{(\text{field})}\,\Delta a \text{ (on line displaced } \boldsymbol{v}\,\Delta b \text{ from start)} \\
+\boldsymbol{u}\,\Delta a, \text{ giving } \quad \nabla_{\boldsymbol{u}}\boldsymbol{A}^{(\text{field})}\,\Delta a \text{ (on line through starting point)} \\
+[\boldsymbol{v},\boldsymbol{u}]\,\Delta a\,\Delta b, \text{ giving } \nabla_{[\boldsymbol{v},\boldsymbol{u}]}\boldsymbol{A}^{(\text{field})}\,\Delta a\,\Delta b
\end{array}\right)
$$

$$
= \{\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{v}} - \nabla_{\boldsymbol{v}}\nabla_{\boldsymbol{u}} + \nabla_{[\boldsymbol{v},\boldsymbol{u}]}\}\boldsymbol{A}^{(\text{field})}\,\Delta a\,\Delta b
$$

$$
= \textbf{\textit{Riemann}}\,(\ldots,\boldsymbol{A}^{(\text{field})},\boldsymbol{u},\boldsymbol{v})\,\Delta a\,\Delta b = \mathcal{R}(\boldsymbol{u},\boldsymbol{v})\boldsymbol{A}^{(\text{field})}\,\Delta a\,\Delta b. \tag{11.15}
$$

Profit: The curvature operator

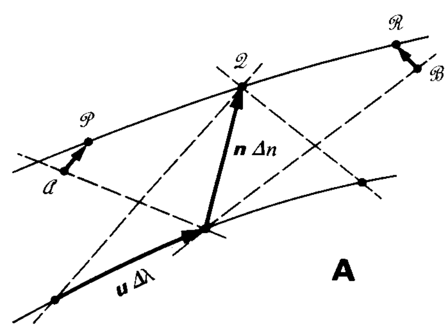
$$
\textbf{\textit{Riemann}}\,(\ldots,\ldots,\boldsymbol{u},\boldsymbol{v}) = \mathcal{R}(\boldsymbol{u},\boldsymbol{v}) = [\nabla_{\boldsymbol{u}},\nabla_{\boldsymbol{v}}] - \nabla_{[\boldsymbol{u},\boldsymbol{v}]},
$$

Box 11.6 GEODESIC DEVIATION AND PARALLEL TRANSPORT AROUND CLOSED CURVE: TWO ASPECTS OF SAME CONSTRUCTION

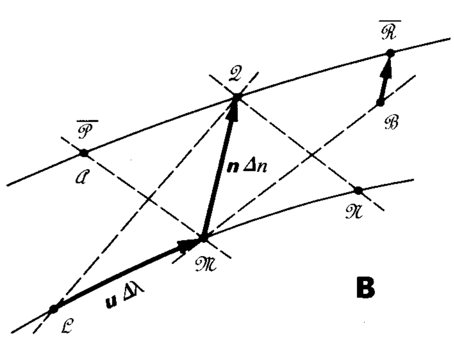
Geodesic Deviation

$$
\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n} = \lim_{\substack{\Delta\lambda\to 0\\ \Delta n\to 0}} \left\{\frac{\mathcal{A}\mathcal{P} + \mathcal{B}\mathcal{R}}{(\Delta\lambda)^2\,\Delta n}\right\}.
$$

(See Boxes 11.2 and 11.3)



A



B

Geodesic Deviation

Same result; different construction. To simplify the connection with closed-curve transport, change the tilt and dilate the parametrization of geodesic $\mathcal{P}\mathcal{Q}\mathcal{R}$ in **A**. The result: **B**, where $\bar{\mathcal{P}}$ and $\mathcal{A}$ coincide. From **F** of Box 11.2 one knows $\mathcal{A}\mathcal{P} + \mathcal{B}\mathcal{R} = \mathcal{A}\bar{\mathcal{P}} + \mathcal{B}\bar{\mathcal{R}}$—i.e. $\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n}$ is the

Box 11.6 ***(continued)***

same for this family of geodesics as for the original family

$$\nabla_u \nabla_u n = \lim_{\substack{\Delta\lambda \to 0 \\ \Delta n \to 0}} \left\{ \frac{\mathcal{B}\bar{\mathcal{R}}}{(\Delta\lambda)^2 \Delta n} \right\}.$$

Also, to simplify discussion set $\Delta n = \Delta\lambda = 1$, and assume $\boldsymbol{n}$ and $\boldsymbol{u}$ are small enough that one can evaluate $\nabla_u \nabla_u n$ without taking the limit:

$$\nabla_u \nabla_u n = \mathcal{B}\bar{\mathcal{R}}.$$

Parallel Transport Around Closed Curve, Performed by Same Construction

Plan: Parallel transport the vector $\boldsymbol{u}\, \Delta\lambda = \mathcal{Q}\bar{\mathcal{R}}$ counterclockwise around the curve $\mathcal{Q} \longrightarrow \bar{\mathcal{P}} \longrightarrow \mathcal{L} \longrightarrow \mathcal{M} \longrightarrow \mathcal{Q}$. *Execution:* (1) Call transported vector $\boldsymbol{u}^{(\mathrm{m})}$ ("m" for "mobile"). (2) At $\mathcal{Q}$, $\boldsymbol{u}^{(\mathrm{m})} = \mathcal{Q}\bar{\mathcal{R}}$. (3) At $\bar{\mathcal{P}}$, $\boldsymbol{u}^{(\mathrm{m})} = \bar{\mathcal{P}}\mathcal{Q}$ because $\bar{\mathcal{P}}\mathcal{Q}\bar{\mathcal{R}}$ is a geodesic and $\boldsymbol{u}^{(\mathrm{m})}$ is its tangent vector. (4) At $\mathcal{L}$, $\boldsymbol{u}^{(\mathrm{m})} = \mathcal{L}\mathcal{M}$ according to Schild's ladder of the picture. (5) At $\mathcal{M}$, $\boldsymbol{u}^{(\mathrm{m})} = \mathcal{M}\mathcal{N}$ because $\mathcal{L}\mathcal{M}\mathcal{N}$ is a geodesic and $\boldsymbol{u}^{(\mathrm{m})}$ is now its tangent vector. (6) At $\mathcal{Q}$, $\boldsymbol{u}^{(\mathrm{m})} = \mathcal{Q}\mathcal{B}$ according to Schild's ladder. Result: The change in $\boldsymbol{u}^{(\mathrm{m})}$ is $-\mathcal{B}\bar{\mathcal{R}}$. Had the curve been circuited in opposite direction ($\mathcal{L} \longrightarrow \bar{\mathcal{P}} \longrightarrow \mathcal{Q} \longrightarrow \mathcal{M} \longrightarrow \mathcal{L}$), the change would have been $+\mathcal{B}\bar{\mathcal{R}}$:

$$(\delta \boldsymbol{u})_{\substack{\text{due to parallel transport up } \boldsymbol{n}, \text{ out } \boldsymbol{u}, \text{ down } -\boldsymbol{n}, \text{ and} \\ \text{back along } -\boldsymbol{u} \text{ to starting point}}} = \mathcal{B}\bar{\mathcal{R}} = \nabla_u \nabla_u n.$$

applied to the vector field $\boldsymbol{A}^{(\mathrm{field})}$, gives the negative of the change in the localized vector $\boldsymbol{A}^{(\mathrm{localized})}$ (called $\boldsymbol{A}^{(\mathrm{mobile})}$ during the phase of travel) on parallel transport around the closed circuit. It does not give the change in $\boldsymbol{A}^{(\mathrm{field})}$ on traversal of that circuit, for $\boldsymbol{A}^{(\mathrm{field})}$ has the same value at the end of the journey as at the beginning. Equation (11.14′) expresses that change in terms of the conveniently calculated differential operator, $\mathcal{R}(\boldsymbol{u}, \boldsymbol{v}) = [\nabla_{\boldsymbol{u}}, \nabla_{\boldsymbol{v}}] - \nabla_{[\boldsymbol{u},\boldsymbol{v}]}$. *Paradox:* Neither wanted nor evaluated is the change in the quantity $\boldsymbol{A}^{(\mathrm{field})}$ acted on by this operator. *Payoff:* Ostensibly differential in the character of its action on $\boldsymbol{A}$, the operator ***Riemann*** $(\ldots, \ldots, \boldsymbol{u}, \boldsymbol{v}) = \mathcal{R}(\boldsymbol{u}, \boldsymbol{v})$ is actually local. Thus, replace the proposed smoothly varying vector field $\boldsymbol{A}^{(\mathrm{field})}$ by a quite different but also smoothly varying vector field $\boldsymbol{A}^{(\mathrm{field,\ new})}$. Then the two fields need agree only at the one point in question for them to give the same output ***Riemann*** $(\ldots, \boldsymbol{A}, \boldsymbol{u}, \boldsymbol{v}) = \mathcal{R}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{A}$ at that point. This one knows from the fact that $\delta\boldsymbol{A}$, the quantity calculated, has an existence and value independent of the choice of $\boldsymbol{A}^{(\text{field})}$. This one can also verify by detailed calculation (exercise 11.2). *Power:* Although they cancel out in their response to any change of $\boldsymbol{A}$ with location, the several differentiations in the curvature operator respond directly to the "rate of change of geometry with location" ("geodesic deviation"). *Prolongation:* The closed curve need not be a quadrilateral. The curvature operator tells how a vector changes on parallel transport about small curves of arbitrary shape (Box 11.7).

EXERCISE

Exercise 11.5. COPLANARITY OF CLOSED CURVES

Let $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ be the bivectors (see Box 11.7) for two small closed curves at the same event. Show that the curves are coplanar if and only if $\boldsymbol{f}_1 = a\boldsymbol{f}_2$ for some number a.

Box 11.7 THE LAW FOR PARALLEL TRANSPORT ABOUT A CLOSED CURVE

A. Special Case

Curve is closed quadrilateral formed by vector fields $\boldsymbol{u}$ and $\boldsymbol{v}$.

1. Law says (in component form)

$$\delta A^\alpha + R^\alpha{}_{\beta\gamma\delta} A^\beta u^\gamma v^\delta = 0. \tag{1}$$

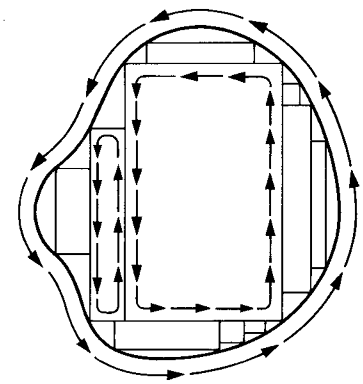
2. On what characteristics of the closed curve does this depend?
   a. Notice that $R^\alpha{}_{\beta\gamma\delta} = -R^\alpha{}_{\beta\delta\gamma}$ (antisymmetry in last two indices; obvious in equation 11.12 for components; also obvious because reversing the direction the curve is traversed—i.e., interchanging $\boldsymbol{u}$ and $\boldsymbol{v}$—should reverse sign of $\delta\boldsymbol{A}$).
   b. Equation (1) contracts $\boldsymbol{u} \otimes \boldsymbol{v}$ into these antisymmetric, last two indices. The symmetric part of $\boldsymbol{u} \otimes \boldsymbol{v}$ must give zero. Only the antisymmetric part, $\boldsymbol{u} \wedge \boldsymbol{v} = \boldsymbol{u} \otimes \boldsymbol{v} - \boldsymbol{v} \otimes \boldsymbol{u}$ can contribute:

$$\delta A^\alpha + \frac{1}{2} R^\alpha{}_{\beta\gamma\delta} A^\beta (\boldsymbol{u} \wedge \boldsymbol{v})^{\gamma\delta} = 0. \tag{2}$$

3. This antisymmetric part is a "*bivector.*" It is independent of the curve's shape; it depends only on (a) the plane the curve lies in, and (b) the area enclosed by the curve. [Although without metric "area" is meaningless, "relative areas at an event in a given plane" have just as much meaning as "relative lengths at an

event along a given direction." Two vectors at the same event lie on the same line if they are multiples of each other; their relative length in that case is their ratio. Similarly, two small closed curves at the same event lie in the same plane if their bivectors are multiples of each other (exercise 11.5); their relative area in that case is the ratio of their bivectors.]

B. General Case



Arbitrary but small closed curve.

1. Break the curve down into a number of quadrilaterals, all lying in the same plane as the curve.
2. Traverse each quadrilateral once in the same sense as the curve is to be traversed. Result: all interior edges get traversed twice in opposite directions (no net traversal); the outer edge (the curve itself) gets traversed once.
3. Thus, $\delta \boldsymbol{A}$ due to traversing curve is the sum of the $\delta \boldsymbol{A}$'s from traversal of each quadrilateral:

$$\delta A^\alpha = -\frac{1}{2} \sum_{\text{quadrilaterals}} R^\alpha{}_{\beta\gamma\delta} A^\beta (\boldsymbol{u} \wedge \boldsymbol{v}_{\text{for given quadrilateral}})^{\gamma\delta}.$$

Define the bivector $\boldsymbol{f}$ for the curve as the sum of the bivectors for its component quadrilaterals:

$$\boldsymbol{f} \equiv \sum_{\text{quadrilaterals}} (\mathbf{u} \wedge \boldsymbol{v})_{\text{quadrilateral}}$$

(add "areas"; keep plane the same).

4. Then

$$\delta A^\alpha + \frac{1}{2} R^\alpha{}_{\beta\gamma\delta} A^\beta f^{\gamma\delta} \equiv \delta A^\alpha + R^\alpha{}_{\beta|\gamma\delta|} A^\beta f^{\gamma\delta} = 0.$$

C. Warning

This is valid only for closed curves of small compass: $\delta \boldsymbol{A}$ doubles when the area doubles; but the error increases by a factor $\sim 2^{3/2}$ [$\delta \boldsymbol{A} \propto \Delta a\, \Delta b$ in calculation of §11.4; but error $\propto (\Delta a)^2\, \Delta b$ or $\Delta a\, (\Delta b)^2$].

§11.5. FLATNESS IS EQUIVALENT TO ZERO RIEMANN CURVATURE

Flatness of a manifold defined

To say that space or spacetime or any other manifold is flat is to say that there exists a coordinate system $\{x^\alpha(\mathscr{P})\}$ in which all geodesics appear straight:

$$x^\alpha(\lambda) = a^\alpha + b^\alpha\lambda. \tag{11.16}$$

(Example: Lorentz spacetime of special relativity, where test bodies move on such straight lines.) They can appear so if and only if the connection coefficients in the geodesic equation

$$\frac{d^2x^\beta}{d\lambda^2} + \Gamma^\beta{}_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0, \tag{11.17}$$

expressed in the same coordinate system, all vanish:

$$\Gamma^\beta{}_{\mu\nu} = 0. \tag{11.18}$$

Flatness implies ***Riemann*** $= 0$

From the vanishing of these connection coefficients, it follows immediately (equation 11.12) that all the components of the curvature tensor are zero:

$$R^\beta{}_{\gamma\mu\nu} = 0. \tag{11.19}$$

[Geometric restatement of (11.16) $\longrightarrow$ (11.18) $\longrightarrow$ (11.19): For all geodesics to be straight in a given coordinate system means that initially parallel geodesics preserve their separation; the geodesic deviation is zero; and therefore the curvature vanishes.]

Is the converse true? Does zero Riemann curvature imply the existence of a coordinate system in which all geodesics appear straight? Yes, as one sees by the following construction.

Proof that ***Riemann*** $= 0$ implies flatness

Transport a vector parallel to itself from $\mathscr{P}_0$ to $\mathscr{Q}$, and then back from $\mathscr{Q}$ to $\mathscr{P}_0$ along a slightly different route. It returns to its starting point with no alteration in magnitude or direction, *because* ***Riemann*** everywhere vanishes. Therefore parallel transport of a base vector $\boldsymbol{e}_\mu$ from $\mathscr{P}_0$ to $\mathscr{Q}$ yields at $\mathscr{Q}$ a base vector $\boldsymbol{e}_\mu$ that is independent, both in magnitude and in direction, of the route of transportation (for routes obtainable one from the other by any continuous sequence of deformations). As for $\mathscr{Q}$, so for all points of the manifold; and as for the one base vector $\boldsymbol{e}_\mu$, so for a complete set of base vectors ($\mu = 0, 1, 2, 3$): Parallel transport of a basis $\{\boldsymbol{e}_\alpha(\mathscr{P}_0)\}$ yields everywhere a field of frames ("frame field"), each base vector of which suffers zero change (relative to the frame field) on parallel transport from any point to any nearby point: thus,

$$\boldsymbol{\nabla}\boldsymbol{e}_\mu = 0; \tag{11.20}$$

or

$$\boldsymbol{\nabla}_\nu\boldsymbol{e}_\mu\,(\equiv \boldsymbol{\nabla}_{\boldsymbol{e}_\nu}\boldsymbol{e}_\mu) = 0. \tag{11.21}$$

With the vanishing of these individual derivatives, there also vanishes the commutator of any two basis-vector fields:

$$[\boldsymbol{e}_\mu, \boldsymbol{e}_\nu] = \boldsymbol{\nabla}_\mu\boldsymbol{e}_\nu - \boldsymbol{\nabla}_\nu\boldsymbol{e}_\mu = 0 - 0 = 0. \tag{11.22}$$

2

The gap in the quadrilateral of Figure 11.2 (there read "$\boldsymbol{e}_\mu$" for "$\boldsymbol{u}$," "$\boldsymbol{e}_\nu$" for "$\boldsymbol{v}$") closes up completely. Thereupon one can introduce coordinates x^μ, each of which increases with a motion in the direction of the corresponding vector field; and with appropriate scaling of these coordinates, one can write

$$\boldsymbol{e}_\mu = \frac{\partial}{\partial x^\mu} \tag{11.23}$$

(see exercise 9.9). With this coordinate basis in hand, one can employ the formula

$$\boldsymbol{\nabla}_\alpha \boldsymbol{e}_\beta = \boldsymbol{e}_\mu \Gamma^\mu{}_{\beta\alpha} \tag{11.24}$$

to calculate the connection coefficients. From the vanishing of the quantities on the left, one concludes that all the connection coefficients on the right ("bending of geodesics") must be zero; so spacetime is indeed flat.

Summary: Spacetime is flat—i.e., there exist "flat coordinates" in which $\Gamma^\mu{}_{\alpha\beta} = 0$ everywhere and geodesics are straight lines, $x^\alpha(\lambda) = a^\alpha + b^\alpha\lambda$—*if and only if* ***Riemann*** $= 0$.

Lorentz coordinates exist if and only if ***Riemann*** $= 0$

Note: In the spacetime of Einstein, which has a metric, one can choose $\{\boldsymbol{e}_\mu(\mathscr{P}_0)\}$ in the above argument to be orthonormal, $\boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu = \eta_{\mu\nu}$ at $\mathscr{P}_0$. The resulting field of frames will then be orthonormal everywhere, and the resulting coordinate system will be Lorentz. Thus, in Einsteinian gravity the above summary can be rewritten: *spacetime is flat (there exists a Lorentz coordinate system) if and only if* ***Riemann*** $= 0$.

Flatness does not imply Euclidean topology

Warning: Flatness does not necessarily imply Euclidean topology. Take a sheet of paper. It is flat. Roll it up into a cylinder. It is still flat, intrinsically. The tracks of geodesics over it have not changed. Distances between neighboring points have not changed. Only the topology has changed, so far as an observer confined forever to the sheet is concerned. (The "extrinsic geometry"—the way the sheet is embedded in the surrounding three-dimensional space—has also changed; but an observer on the sheet knows nothing of this, and it is not the subject of the present chapter. See, instead §21.5.)

Take this cylinder. Bend it around and glue its two ends together, without changing its flat intrinsic geometry. Doing so is impossible if the cylinder remains embedded in flat, three-dimensional Euclidean space; perfectly possible if it is embedded in a Euclidean space of 4 dimensions. However, embedding is unimportant to observers confined to the cylinder, since all they ever measure is intrinsic geometry; so all that matters to them is the *topological identification* of the two ends of the cylinder with each other. The result is topologically a torus; but the tracks of geodesics are still unchanged; the intrinsic geometry is flat; ***Riemann*** vanishes.

By analogy, take flat Minkowskii spacetime. Pick some Lorentz frame, and in it pick a cube 10^{10} light years on each side ($0 < x < 10^{10}$ light years; similarly for y and z). Identify opposite faces of the cube so that a geodesic exiting across one face enters across the other. The result is topologically a three-torus: a "closed universe" with finite volume, with flat, Minkowskii geometry, and with a form that changes not at all as Lorentz time t passes (no expansion, no contraction).

§11.6. RIEMANN NORMAL COORDINATES

In curved spacetime one can never find a coordinate system with $\Gamma^{\alpha}{}_{\beta\gamma} = 0$ everywhere. But one can always construct local inertial frames at a given event $\mathscr{P}_0$; and as viewed in such frames, free particles must move along straight lines, at least locally—which means $\Gamma^{\alpha}{}_{\beta\gamma}$ must vanish, at least locally.

Riemann normal coordinates: a realization of local inertial frames

A very special and useful realization of such a local inertial frame is a *Riemann-normal coordinate system*. Pick an event $\mathscr{P}_0$ and a set of basis vectors $\{\boldsymbol{e}_\alpha(\mathscr{P}_0)\}$ to be used there by an inertial observer. Fill spacetime, near $\mathscr{P}_0$, with geodesics radiating out from $\mathscr{P}_0$ like the quills of a hedgehog or porcupine. Each geodesic is determined by its tangent vector $\boldsymbol{v}$ at $\mathscr{P}_0$; and the general point on it can be denoted

Geometric construction of Riemann normal coordinates

$$\mathscr{P} = \mathscr{G}(\underbrace{\lambda}_{\text{affine parameter; tells "where" on geodesic}}; \underbrace{\boldsymbol{v}}_{\text{tangent vector at } \mathscr{P}_0\text{; tells "which geodesic"}}). \tag{11.25}$$

Actually, this gives more geodesics than are needed. One reaches the same point after parameter length $\frac{1}{2}\lambda$ if the initial tangent vector is $2\boldsymbol{v}$, as one reaches after λ if the tangent vector is $\boldsymbol{v}$:

$$\mathscr{G}(\lambda; \boldsymbol{v}) = \mathscr{G}\left(\frac{1}{2}\lambda; 2\boldsymbol{v}\right) = \mathscr{G}(1; \lambda\boldsymbol{v}).$$

Thus, by fixing $\lambda = 1$ and varying $\boldsymbol{v}$ in all possible ways, one can reach every point in some neighborhood of $\mathscr{P}_0$. This is the foundation for constructing Riemann normal coordinates. Choose an event $\mathscr{P}$. Find that tangent vector $\boldsymbol{v}$ at $\mathscr{P}_0$ for which $\mathscr{P} = \mathscr{G}(1; \boldsymbol{v})$. Expand that $\boldsymbol{v}$ in terms of the chosen basis and give its components the names x^α:

$$\mathscr{P} = \mathscr{G}(1; x^\alpha \boldsymbol{e}_\alpha). \tag{11.26}$$

The point $\mathscr{P}$ determines x^α uniquely (if $\mathscr{P}$ is near enough to $\mathscr{P}_0$ that spacetime curvature has not caused geodesics to cross each other). Similarly, x^α determines $\mathscr{P}$ uniquely. Hence, x^α can be chosen as the coordinates of $\mathscr{P}$—its "Riemann-normal coordinates, based on the event $\mathscr{P}_0$ and basis $\{\boldsymbol{e}_\alpha(\mathscr{P}_0)\}$."

Mathematical properties of Riemann normal coordinates

Equation (11.26) summarizes Riemann-normal coordinates concisely. Other equations, derived in exercise 11.9, summarize their powerful properties:

$$\boldsymbol{e}_\alpha(\mathscr{P}_0) = (\partial/\partial x^\alpha)_{\mathscr{P}_0}; \tag{11.27}$$

$$\Gamma^{\alpha}{}_{\beta\gamma}(\mathscr{P}_0) = 0; \tag{11.28}$$

$$\Gamma^{\alpha}{}_{\beta\gamma,\mu}(\mathscr{P}_0) = -\frac{1}{3}(R^{\alpha}{}_{\beta\gamma\mu} + R^{\alpha}{}_{\gamma\beta\mu}). \tag{11.29}$$

If spacetime has a metric (as it does in actuality), and if the observer's frame at $\mathscr{P}_0$ has been chosen orthonormal ($\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta = \eta_{\alpha\beta}$), then

$$g_{\alpha\beta}(\mathscr{P}_0) = \eta_{\alpha\beta}, \tag{11.30}$$

$$g_{\alpha\beta,\mu}(\mathscr{P}_0) = 0, \tag{11.31}$$

$$\begin{aligned} g_{\alpha\beta,\mu\nu}(\mathscr{P}_0) &= -\frac{1}{3}(R_{\alpha\mu\beta\nu} + R_{\alpha\nu\beta\mu}) \\ &= -\frac{2}{3}J_{\alpha\beta\mu\nu}, \end{aligned} \tag{11.32}$$

$$R_{\alpha\beta\gamma\delta}(\mathscr{P}_0) = g_{\alpha\delta,\beta\gamma}(\mathscr{P}_0) - g_{\alpha\gamma,\beta\delta}(\mathscr{P}_0). \tag{11.32'}$$

Here $J_{\alpha\beta\mu\nu}$ are components of the Jacobi curvature tensor (see exercise 11.7).

Other mathematical realizations of a local inertial frame

Is this the only coordinate system that is locally inertial at $\mathscr{P}_0$ (i.e., has $\Gamma^\alpha{}_{\beta\gamma} = 0$ there) and is tied to the basis vectors $\boldsymbol{e}_\alpha$ there (i.e., has $\partial/\partial x^\alpha = \boldsymbol{e}_\alpha$ there)? No. But all such coordinate systems (called "*normal coordinates*") will be the same to second order:

$$x^\alpha_{\text{NEW}}(\mathscr{P}) = x^\alpha_{\text{OLD}}(\mathscr{P}) + \text{corrections of order } (x^\alpha_{\text{OLD}})^3.$$

Moreover, only those the same to third order,

$$x^\alpha_{\text{NEW}}(\mathscr{P}) = x^\alpha_{\text{OLD}}(\mathscr{P}) + \text{corrections of order } (x^\alpha_{\text{OLD}})^4,$$

will preserve the beautiful ties (11.29) and (11.32) to the Riemann curvature tensor.

EXERCISES

Exercise 11.6. SYMMETRIES OF *Riemann*

(To be discussed in Chapter 13). Show that ***Riemann*** has the following symmetries:

$$R^\alpha{}_{\beta\gamma\delta} = R^\alpha{}_{\beta[\gamma\delta]} \quad \text{(antisymmetric on last 2 indices)} \tag{11.33a}$$

$$R^\alpha{}_{[\beta\gamma\delta]} = 0 \quad \text{(vanishing of completely antisymmetric part)} \tag{11.33b}$$

Exercise 11.7. GEODESIC DEVIATION MEASURES ALL CURVATURE COMPONENTS

The equation of geodesic deviation, written up to now in the form

$$\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n} + \textbf{\textit{Riemann}}(\ldots, \boldsymbol{u}, \boldsymbol{n}, \boldsymbol{u}) = 0$$

or

$$\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n} + \mathscr{R}(\boldsymbol{n}, \boldsymbol{u})\boldsymbol{u} = 0,$$

also lets itself be written in the Jacobi form $\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n} + \mathscr{J}(\boldsymbol{u}, \boldsymbol{u})\boldsymbol{n} = 0$. Here $\mathscr{J}(\boldsymbol{u}, \boldsymbol{v})$, the "*Jacobi curvature operator,*" is defined by

$$\mathscr{J}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{n} \equiv \frac{1}{2}[\mathscr{R}(\boldsymbol{n}, \boldsymbol{u})\boldsymbol{v} + \mathscr{R}(\boldsymbol{n}, \boldsymbol{v})\boldsymbol{u}], \tag{11.34}$$

and is related to the "Jacobi curvature tensor" by

$$\textbf{\textit{Jacobi}}(\ldots, \boldsymbol{n}, \boldsymbol{u}, \boldsymbol{v}) \equiv \mathscr{J}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{n}, \tag{11.35}$$

which implies

$$J^{\mu}{}_{\nu\alpha\beta} = J^{\mu}{}_{\nu\beta\alpha} = \frac{1}{2}(R^{\mu}{}_{\alpha\nu\beta} + R^{\mu}{}_{\beta\nu\alpha}). \tag{11.36}$$

(a) Show that $J^{\mu}{}_{(\alpha\beta\gamma)} = 0$ follows from $R^{\mu}{}_{\alpha\beta\gamma} = R^{\mu}{}_{\alpha[\beta\gamma]}$.

(b) Show that by studying geodesic deviation (allowing arbitrary $\boldsymbol{u}$ and $\boldsymbol{n}$ in $\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{n} + \mathcal{J}(\boldsymbol{u}, \boldsymbol{u})\boldsymbol{n} = 0$) one can measure *all* components of ***Jacobi***.

(c) Show that ***Jacobi*** contains precisely the same information as ***Riemann***. [Hint: show that

$$R^{\mu}{}_{\alpha\nu\beta} = \frac{2}{3}(J^{\mu}{}_{\nu\alpha\beta} - J^{\mu}{}_{\beta\alpha\nu}); \tag{11.37}$$

this plus equation (11.36) for $J^{\mu}{}_{\nu\alpha\beta}$ proves "same information content".] Hence, by studying geodesic deviation one can also measure all the components of ***Riemann***.

(d) Show that the symmetry of $R^{\mu}{}_{[\nu\alpha\beta]} = 0$ is essential in the equivalence between ***Jacobi*** and ***Riemann*** by exhibiting proposed values for $R^{\mu}{}_{\nu\alpha\beta} = -R^{\mu}{}_{\nu\beta\alpha}$ for which $R^{\mu}{}_{[\nu\alpha\beta]} \neq 0$, and from which one would find $J^{\mu}{}_{\nu\alpha\beta} = 0$.

Exercise 11.8. GEODESIC DEVIATION IN GORY DETAIL

Write out the equation of geodesic deviation in component form in a coordinate system. Expand all covariant derivatives (semicolon notation) in terms of ordinary (comma) derivatives and in terms of Γ's to show all Γ and ∂ terms explicitly.

Exercise 11.9. RIEMANN NORMAL COORDINATES IN GENERAL

Derive properties (11.27), (11.28), (11.29), (11.31), (11.32), and (11.32′) of Riemann normal coordinates. *Hint:* Proceed as follows.

(a) From definition (11.26), derive $(\partial\mathcal{P}/\partial x^{\alpha})_{\mathcal{P}_0} = \boldsymbol{e}_{\alpha}$.

(b) Similarly, from definition (11.26), show that each of the curves $x^{\alpha} = v^{\alpha}\lambda$ (where the v^{α} are constants) is a geodesic through $\mathcal{P}_0$, with affine parameter λ.

(c) Show that $\Gamma^{\alpha}{}_{\beta\gamma}(\mathcal{P}_0) = 0$ by substituting $x^{\alpha} = v^{\alpha}\lambda$ into the geodesic equation.

(d) Since the curves $x^{\alpha} = v^{\alpha}\lambda$ are geodesics for every choice of the parameters v^{α}, they provide not only a geodesic tangent $\boldsymbol{u} \equiv (\partial/\partial\lambda)_{v^{\alpha}}$, but also several deviation vectors $\boldsymbol{N}_{(\alpha)} \equiv (\partial/\partial v^{\alpha})_{\lambda}$. Compute the components of these vectors in the Riemann normal coordinate system, and substitute into the geodesic deviation equation as written in exercise 11.8.

(e) Equate to zero the coefficients of the zeroth and first powers of λ in the geodesic deviation equation of part (d), using

$$\Gamma^{\alpha}{}_{\beta\gamma}\Big|_{x^{\mu}=v^{\mu}\lambda} = \lambda v^{\mu}\Gamma^{\alpha}{}_{\beta\gamma,\mu}(\mathcal{P}_0) + O(\lambda^2),$$

which is a Taylor series for Γ. In this way arrive at equation (11.29) for $\Gamma^{\alpha}{}_{\beta\gamma,\mu}$ in terms of the Riemann tensor.

(f) From equations (11.28), (11.29), and (8.24) for the connection coefficients in terms of the metric, derive equations (11.31), (11.32), and (11.32′).

Exercise 11.10. BIANCHI IDENTITIES

Show that the Riemann curvature tensor satisfies the following "Bianchi identities"

$$R^{\alpha}{}_{\beta[\gamma\delta;\epsilon]} = 0. \tag{11.38}$$

The geometric meaning of these identities will be discussed in Chapter 15. [*Hint:* Perform the calculation at the origin of a Riemann normal coordinate system.]

Exercise 11.11. CURVATURE OPERATOR ACTS ON 1-FORMS

Let $\mathcal{R}(\boldsymbol{u}, \boldsymbol{v})$ be the operator $\mathcal{R}(\boldsymbol{u}, \boldsymbol{v}) = [\nabla_{\boldsymbol{u}}, \nabla_{\boldsymbol{v}}] - \nabla_{[\boldsymbol{u}, \boldsymbol{v}]}$ when acting on 1-forms $\boldsymbol{\sigma}$ (or other tensors) as well as on tangent vectors. Show that

$$\langle \mathcal{R}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{\sigma}, \boldsymbol{w}\rangle = -\langle \boldsymbol{\sigma}, \mathcal{R}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{w}\rangle.$$

Exercise 11.12. ROTATION GROUP: RIEMANN CURVATURE

[Continuation of exercises 9.13, 9.14, and 10.17.] Calculate the components of the Riemann curvature tensor for the rotation group's manifold $SO(3)$; use the basis of generators $\{\boldsymbol{e}_\alpha\}$. [*Answer:*

$$R^\alpha{}_{\beta\gamma\delta} = \frac{1}{2}\delta^{\alpha\beta}_{\gamma\delta}, \tag{11.39}$$

where $\delta^{\alpha\beta}_{\gamma\delta}$ is the permutation symbol defined in equation (3.50l):

$$\delta^{\alpha\beta}_{\gamma\delta} \equiv (\delta^\alpha{}_\gamma \delta^\beta{}_\delta - \delta^\alpha{}_\delta \delta^\beta{}_\gamma).$$

Note that this answer is independent of location $\mathcal{P}$ in the group manifold.]

CHAPTER 12

NEWTONIAN GRAVITY IN THE LANGUAGE OF CURVED SPACETIME

The longest period of time for which a modern painting has hung upside down in a public gallery unnoticed is 47 days. This occurred to Le Bateau *by Matisse in the Museum of Modern Art, New York City. In this time 116,000 people had passed through the gallery.*

McWHIRTER AND McWHIRTER (1971)

§12.1. NEWTONIAN GRAVITY IN BRIEF

This chapter is entirely Track 2. Chapters 9–11 are necessary preparation for it. It is not needed for any later chapter, but it will be helpful in
(1) Chapter 17 (Einstein field equations) and
(2) Chapters 38 and 39 (experimental tests and other theories of gravity).

The equivalence principle is not unique to Einstein's description of the facts of gravity. What is unique to Einstein is the combination of the equivalence principle and local Lorentz geometry. To return to the world of Newton, forget everything discovered in the last century about special relativity, light cones, the limiting speed of light, and proper time. Return to the "universal time" t of earlier centuries. In terms of that universal time, and of rectangular, "Galilean" space coordinates, Newtonian theory gives for the trajectories of neutral test particles

Newtonian gravity: original formulation

$$\frac{d^2x^j}{dt^2} + \frac{\partial\Phi}{\partial x^j} = 0; \tag{12.1}$$

$$\Phi \text{ (sometimes denoted } -U) = \text{Newtonian potential.} \tag{12.2}$$

Customarily one interprets these equations as describing the "curved paths" $x^j(t)$ along which test particles fall in Euclidean space (*not* spacetime). These curved paths include circular orbits about the Earth and the parabolic trajectory of a baseball. Cartan (1923, 1924) asks one to abandon this viewpoint. Instead, he says, regard these trajectories as geodesics $[t(\lambda), x^j(\lambda)]$ in curved spacetime. (This change of viewpoint was embodied in Figures B and C of Box 1.6.) Since the "affinely ticking"

Newtonian gravity: translation into language of curved spacetime

Newtonian clocks carried by test particles read universal time (or some multiple. $\lambda = at + b$, thereof), the equation of motion (12.1) can be rewritten

$$\frac{d^2t}{d\lambda^2} = 0, \qquad \frac{d^2x^j}{d\lambda^2} + \frac{\partial\Phi}{\partial x^j}\left(\frac{dt}{d\lambda}\right)^2 = 0. \tag{12.3}$$

By comparing with the geodesic equation

$$d^2x^\alpha/d\lambda^2 + \Gamma^\alpha{}_{\beta\gamma}(dx^\beta/d\lambda)(dx^\gamma/d\lambda) = 0,$$

one can read off the values of the connection coefficients:

$$\Gamma^j{}_{00} = \partial\Phi/\partial x^j; \qquad \text{all other } \Gamma^\alpha{}_{\beta\gamma} \text{ vanish.} \tag{12.4}$$

And by inserting these into the standard equation (11.12) for the components of the Riemann tensor, one learns (exercise 12.1)

$$R^j{}_{0k0} = -R^j{}_{00k} = \frac{\partial^2\Phi}{\partial x^j\,\partial x^k}; \qquad \text{all other } R^\alpha{}_{\beta\gamma\delta} \text{ vanish.} \tag{12.5}$$

Finally, the source equation for the Newtonian potential

$$\nabla^2\Phi \equiv \sum_j \Phi_{,jj} = 4\pi\rho \tag{12.6}$$

one can rewrite with the help of the *"Ricci curvature tensor"*

$$R_{\alpha\beta} \equiv R^\mu{}_{\alpha\mu\beta} \text{ (contraction of } \textbf{\textit{Riemann}}) \tag{12.7}$$

in the geometric form (exercise 12.2)

$$R_{00} = 4\pi\rho; \qquad \text{all other } R_{\alpha\beta} \text{ vanish.} \tag{12.8}$$

Equation (12.4) for $\Gamma^\alpha{}_{\beta\gamma}$, equation (12.5) for $R^\alpha{}_{\beta\gamma\delta}$, equation (12.8) for $R_{\alpha\beta}$, plus the law of geodesic motion are the full content of Newtonian gravity, rewritten in geometric language.

It is one thing to pass quickly through these component manipulations. It is quite another to understand fully, in abstract and pictorial terms, the meanings of these equations and the structure of Newtonian spacetime. To produce such understanding, and to compare Newtonian spacetime with Einsteinian spacetime, are the goals of this chapter, which is based on the work of Cartan (1923, 1924), Trautman (1965), and Misner (1969a).

EXERCISES

Exercise 12.1. RIEMANN CURVATURE OF NEWTONIAN SPACETIME

Derive equation (12.5) for $R^\alpha{}_{\beta\gamma\delta}$ from equation (12.4) for $\Gamma^\alpha{}_{\beta\gamma}$.

Exercise 12.2. NEWTONIAN FIELD EQUATION

Derive the geometric form (12.8) of the Newtonian field equation from (12.5) through (12.7).

---

§12.2. STRATIFICATION OF NEWTONIAN SPACETIME

Galileo and Newton spoke of a flat, Euclidean "absolute space" and of an "absolute time," two concepts distinct and unlinked. In absolute space Newtonian physics took place; and as it took place, absolute time marched on. No hint was there that space and time might be two aspects of a single entity, a curved *"spacetime"*—until Einstein made the unification in relativity physics, and Cartan (1923) followed suit in Newtonian physics in order to provide clearer insight into Einstein's ideas.

The geometry of Newtonian spacetime:

How do the absolute space of Galileo and Newton, and their absolute time, fit into Cartan's *"Newtonian spacetime"?* The key to the fit is *stratification;* stratification produced by the universal time coordinate t.

"Universal time" as a scalar field

Regard t as a function (scalar field) defined once and for all in Newtonian spacetime

$$t = t(\mathscr{P}). \tag{12.9}$$

Without it, spacetime could not be Newtonian, for "t" is every bit as intrinsic to Newtonian spacetime as the metric "$\boldsymbol{g}$" is to Lorentz spacetime. The layers of spacetime are the slices of constant t—the *"space slices"*—each of which has an identical geometric structure: the old "absolute space."

Adopting Cartan's viewpoint, ask what kind of geometry is induced onto each space slice by the surrounding geometry of spacetime. A given space slice is endowed, by the Galilean coordinates of §12.1, with basis vectors $\boldsymbol{e}_j = \partial/\partial x^j$; and this basis has vanishing connection coefficients, $\Gamma^j{}_{k\ell} = 0$ [cf. equation (12.4)]. Consequently, *the geometry of each space slice is completely flat.*

Space slices with Euclidean geometry

"Absolute space" is Euclidean in its geometry, according to the old viewpoint, and the Galilean coordinates are Cartesian. Translated into Cartan's language, this says: not only is each space slice ($t =$ constant) flat, and not only do its Galilean coordinates have vanishing connection coefficients, but also *each space slice is endowed with a three-dimensional metric, and its Galilean coordinate basis is orthonormal,*

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = (\partial/\partial x^i)\cdot(\partial/\partial x^j) = \delta_{ij}. \tag{12.10}$$

Curvature acts in spacetime, not in space slices

If the space slices are really so flat, where do curvature and geodesic deviation enter in? They are properties of *spacetime.* Parallel transport a vector around a closed curve lying entirely in a space slice; it will return to its starting point unchanged. But transport it forward in time by Δt, northerly in space by Δx^k, back in time by $-\Delta t$, and southerly by $-\Delta x^k$ to its starting point; it will return changed by

$$\delta\boldsymbol{A} = -\mathscr{R}\left(\Delta t\,\frac{\partial}{\partial t},\ \Delta x^k\,\frac{\partial}{\partial x^k}\right)\boldsymbol{A};$$

i.e.,

$$\delta A^0 = 0, \qquad \delta A^j = -R^j{}_{00k}A^0(\Delta t)(\Delta x^k) = \frac{\partial^2\Phi}{\partial x^j\,\partial x^k}A^0(\Delta t)(\Delta x^k). \tag{12.11}$$

Geodesics of a space slice (Euclid's straight lines) that are initially parallel remain always parallel. But geodesics of spacetime (trajectories of freely falling particles) initially parallel get pried apart or pushed together by spacetime curvature,

$$\nabla_{\boldsymbol{u}} \nabla_{\boldsymbol{u}} \boldsymbol{n} + \mathcal{R}(\boldsymbol{n}, \boldsymbol{u})\boldsymbol{u} = 0,$$

or equivalently in Galilean coordinates:

$$n^0 = dn^0/dt = 0 \text{ initially} \Longrightarrow n^0 = 0 \text{ always}; \tag{12.12a}$$

$$\frac{d^2 n^j}{dt^2} + \frac{\partial^2 \Phi}{\partial x^j \, \partial x^k} n^k = 0 \tag{12.12b}$$

(see Box 12.1 and exercise 12.3).

EXERCISE

Exercise 12.3. GEODESIC DEVIATION DERIVED

Produce a third column for Box 11.4, one that carries out the "geometric analysis" in component notation using the Galilean connection coefficients (12.4) of Newtonian spacetime. Thereby achieve a deeper understanding of how the geometric analysis parallels the old Newtonian analysis.

---

§12.3. GALILEAN COORDINATE SYSTEMS

The Lorentz spacetime of special relativity has an existence and structure completely independent of any coordinate system. But a special property of its geometry (zero curvature) allows the introduction of a special class of coordinates (Lorentz coordinates), which cling to spacetime in a special way

$$(\partial/\partial x^\alpha) \cdot (\partial/\partial x^\beta) = \eta_{\alpha\beta} \text{ everywhere.}$$

By studying these special coordinate systems and the relationships between them (Lorentz transformations), one learns much about the structure of spacetime itself (breakdown in simultaneity; Lorentz contraction; time dilatation; . . .).

Galilean coordinates defined

Similarly for Newtonian spacetime. Special properties of its geometry (explored in abstract later; Box 12.4) permit the introduction of special coordinates (Galilean coordinates), which cling to spacetime in a special way

$$x^0(\mathcal{P}) = t(\mathcal{P});$$

$$(\partial/\partial x^j) \cdot (\partial/\partial x^k) = \delta_{jk};$$

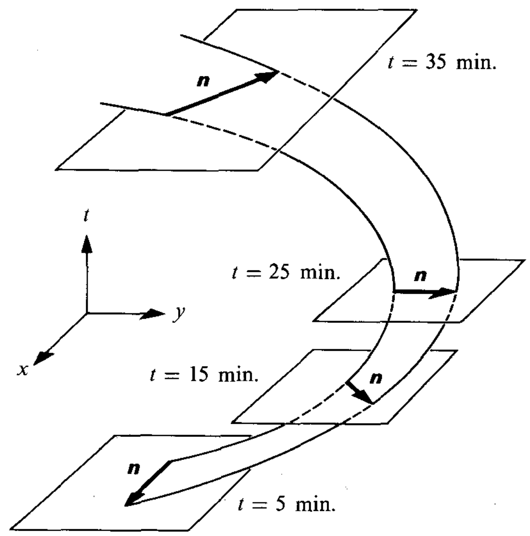
$$\Gamma^j{}_{00} = \Phi_{,j} \text{ for some scalar field } \Phi, \text{ and all other } \Gamma^\alpha{}_{\beta\gamma} \text{ vanish.}$$

To understand Newtonian spacetime more deeply, study the relations between these Galilean coordinate systems.

Box 12.1 GEODESIC DEVIATION IN NEWTONIAN SPACETIME

Coordinate system for calculation: Galilean space coordinates x^j and universal time coordinate t.
General component form of equation:

$$\frac{D^2 n^\alpha}{d\lambda^2} + R^\alpha{}_{\beta\gamma\delta}\frac{dx^\beta}{d\lambda} n^\gamma \frac{dx^\delta}{d\lambda} = 0.$$



Special conditions for this calculation: let the particles' clocks (affine parameters) all be normalized to read universal time, $\lambda = t$. This means that the separation vector

$$n^\alpha = (\partial x^\alpha/\partial n)_\lambda$$

between geodesics has zero time component, $n^0 = 0$; i.e., in abstract language,

$$\langle \boldsymbol{d}t, \boldsymbol{n}\rangle = t_{,\alpha} n^\alpha = n^0 = 0;$$

i.e., in geometric language, $\boldsymbol{n}$ lies in a space slice (surface of constant t).
Evaluation of covariant derivative:

$$\frac{Dn^\alpha}{D\lambda} = \frac{dn^\alpha}{d\lambda} + \underbrace{\Gamma^\alpha{}_{\beta\mu}}_{[0 \text{ unless } \beta = 0]} \underset{[0 \text{ unless } \beta \text{ is a space index } (n^0 = 0)]}{n^\beta} \frac{dx^\mu}{d\lambda} = \frac{dn^\alpha}{d\lambda},$$

$$\frac{D^2 n^\alpha}{d\lambda^2} = \frac{d(Dn^\alpha/d\lambda)}{d\lambda} + \underbrace{\Gamma^\alpha{}_{\beta\mu}}_{[0 \text{ unless } \beta = 0]} \underset{[0 \text{ unless } \beta \text{ is space index}]}{\frac{dn^\beta}{d\lambda}} \frac{dx^\mu}{d\lambda} = \frac{d(Dn^\alpha/d\lambda)}{d\lambda} = \frac{d^2 n^\alpha}{d\lambda^2} \underset{[\text{since } \lambda = t]}{=} \frac{d^2 n^\alpha}{dt^2}.$$

Evaluation of tidal accelerations:

$$R^0{}_{\beta\gamma\delta}\frac{dx^\beta}{d\lambda} n^\gamma \frac{dx^\delta}{d\lambda} = 0 \qquad \text{since } R^j{}_{0k0} \text{ and } R^j{}_{00k} \text{ are only nonzero components.}$$

$$\underbrace{R^j{}_{\beta\gamma\delta}}_{[\text{for } \gamma \text{ a space index: } 0 \text{ unless } \beta = \delta = 0]} \frac{dx^\beta}{d\lambda} \underset{[0 \text{ unless } \gamma \text{ is space index}]}{n^\gamma} \frac{dx^\delta}{d\lambda} = R^j{}_{0k0}\frac{dt}{d\lambda} n^k \frac{dt}{d\lambda} = R^j{}_{0k0} n^k = \frac{\partial^2 \Phi}{\partial x^j \partial x^k} n^k.$$

Resultant equation of geodesic deviation:

$$\frac{d^2 n^0}{dt^2} = 0 \qquad \left(\begin{array}{l}\text{agrees with result } n^0 = 0 \text{ always, which} \\ \text{followed from choice } \lambda = t \text{ for all particles}\end{array}\right)$$

$$\frac{d^2 n^j}{dt^2} + \frac{\partial^2 \Phi}{\partial x^j \partial x^k} n^k = 0 \qquad \left(\begin{array}{l}\text{agrees with Newton-type calculation} \\ \text{in Box 11.4; see also exercise 12.3}\end{array}\right).$$

Point of principle: how can one write down the laws of gravity and properties of spacetime in Galilean coordinates first (§12.1), and only afterward (here) come to grip with the nature of the coordinate system and its nonuniqueness? Answer: (a quotation from §3.1, slightly modified): "Here and elsewhere in science, as emphasized not least by Henri Poincaré, that view is out of date which used to say 'Define your terms before you proceed.' All the laws and theories of physics, including Newton's laws of gravity, have this deep and subtle character, that they both define the concepts they use (here Galilean coordinates) and make statements about these concepts."

The Newtonian laws of gravity, written in a Galilean coordinate system

$$x^0 = t, \qquad (\partial/\partial x^j)\cdot(\partial/\partial x^k) = \delta_{jk}$$

make the statement "$\Gamma^j{}_{00} = \Phi_{,j}$ and all other $\Gamma^\alpha{}_{\beta\gamma} = 0$" about the geometry of spacetime. This statement in turn gives information about the relationships between different Galilean systems. Let one Galilean system $\{x^\alpha(\mathscr{P})\}$ be given, and seek the most general coordinate transformation leading to another, $\{x^{\alpha'}(\mathscr{P})\}$. The following constraints exist: (1) $x^{0'} = x^0 = t$ (both time coordinates must be universal time); (2) at fixed t (i.e., in a fixed space slice) both sets of space coordinates must be Euclidean, so they must be related by a rotation and a translation:

$$x^{j'} = \underbrace{A_{j'k}}_{[\text{rotation matrix, i.e., } A_{j'\ell}A_{k'\ell} = \delta_{j'k'}]} x^k + \underbrace{a^{j'}}_{[\text{translation}]} \tag{12.13a}$$

$$x^k = A_{j'k}x^{j'} - a^k, \text{ with } a^k \equiv A_{j'k}a^{j'}. \tag{12.13b}$$

The rotation and translation might, *a priori*, be different on different slices, $A_{j'k} = A_{j'k}(t)$ and $a^j = a^j(t)$; but (3) they must be constrained by the required special form of the connection coefficients. Calculate the connection coefficients in the new coordinate system, given their form in the old. The result (exercise 12.4) is:

$$\begin{aligned}
&\Gamma^{j'}{}_{0'k'} = \Gamma^{j'}{}_{k'0'} = A_{j'\ell}\dot{A}_{k'\ell} \qquad (\text{produces "Coriolis forces"});\\
&\Gamma^{j'}{}_{0'0'} = \frac{\partial\Phi}{\partial x^{j'}} + A_{j'k}(\underbrace{\ddot{A}_{\ell'k}x^{\ell'}}_{[\text{"centrifugal forces"}]} - \underbrace{\ddot{a}^k}_{[\text{"inertial forces"}]});\\
&\text{all other } \Gamma^{\alpha'}{}_{\beta'\gamma'} \text{ vanish}
\end{aligned} \tag{12.14}$$

("Euclidean" index conventions; repeated space indices to be summed even if both are down; dot denotes time derivative). These have the standard Galilean form (12.4) if and only if

$$\dot{A}_{j'k} = 0, \qquad \underbrace{\Phi'}_{\left[\substack{\text{Newtonian potential in}\\ \text{new coordinate system}}\right]} = \underbrace{\Phi}_{\left[\substack{\text{Newtonian potential in}\\ \text{old coordinate system}}\right]} - \ddot{a}^k x^k + \text{constant}. \tag{12.15}$$

These results can be restated in words: any two Galilean coordinate systems are related by (1) a time-independent rotation of the space grid (*same* rotation on each space slice), and (2) a time-dependent translation of the space grid (translation possibly *different* on different slices)

Transformations linking Galilean coordinate systems

$$x^{j'} = \underset{[\text{constant}]}{A_{j'k}}\, x^k + \underset{[\text{time-dependent}]}{a^{j'}(t)}. \tag{12.16}$$

The Newtonian potential is not a function defined in spacetime with existence independent of all coordinate systems. (There is no coordinate-free way to measure it.) Rather, it depends for its existence on a particular choice of Galilean coordinates; and if the choice is changed via equation (12.16), then Φ is changed:

Newtonian potential depends on choice of Galilean coordinate system

$$\Phi' = \Phi - \ddot{a}^k x^k. \tag{12.17}$$

(By contrast, an existence independent of all coordinates is granted to the universal time $t(\mathscr{P})$ and the covariant derivative $\boldsymbol{\nabla}$.)

Were all the matter in the universe concentrated in a finite region of space and surrounded by emptiness ("island universe"), then one could impose the global boundary condition

Absolute Galilean coordinates defined

$$\Phi \longrightarrow 0 \text{ as } r \equiv (x^k x^k)^{1/2} \longrightarrow \infty. \tag{12.18}$$

This would single out a subclass of Galilean coordinates ("*absolute*" Galilean coordinates), with a unique, common Newtonian potential. The transformation from one absolute Galilean coordinate system to any other would be

Transformations linking absolute Galilean coordinate systems

$$x^{j'} = \underset{\left[\substack{\text{constant}\\ \text{rotation}}\right]}{A_{j'k}}\, x^k + \underset{\left[\substack{\text{constant}\\ \text{displacement}}\right]}{a^{j'}} + \underset{\left[\substack{\text{constant}\\ \text{velocity}}\right]}{v^{j'}}\, t \tag{12.19}$$

("Galilean transformation"). *But,* (1) by no local measurements could one ever distinguish these absolute Galilean coordinate systems from the broader class of Galilean systems (to distinguish, one must integrate the locally measurable quantity $\Phi_{,j} = \Gamma^j{}_{00}$ out to infinity); and (2) astronomical data deny that the real universe is an island of matter surrounded by emptiness.

It is instructive to compare Galilean coordinates and Newtonian spacetime as described above with Lorentz coordinates and the Minkowskii spacetime of special relativity, and with the general coordinates and Einstein spacetime of general relativity; see Boxes 12.2 and 12.3.

(*continued on page 298*)

2

Box 12.2 NEWTONIAN SPACETIME, MINKOWSKIIAN SPACETIME, AND EINSTEINIAN SPACETIME: COMPARISON AND CONTRAST

Query	*Newtonian spacetime*	*Minkowskiian spacetime (special relativity)*	*Einsteinian spacetime (general relativity)*
What *a priori* geometric structures does spacetime possess?	(1) Universal time function t (2) Covariant derivative ∇ (3) Spatial metric "$\cdot$"; but spacetime metric can *not* be defined (exercise 12.10)	A spacetime metric that is flat (vanishing Riemann curvature)	A spacetime metric
What preferred coordinate systems are present?	(1) Galilean coordinates in general (2) Absolute Galilean coordinates in an island universe (this case not considered here)	Lorentz coordinates	In general, every coordinate system is equally preferred (though in special cases with symmetry there are special preferred coordinates)
What is required to select out a particular preferred coordinate system?	(1) A single spatial orientation, the same throughout all spacetime (three Euler angles) (2) The arbitrary world line of the origin of space coordinates (three functions of time)	(1) A single spatial orientation, the same throughout all spacetime (three Euler angles) (2) The location of the origin of coordinates (four numbers) (3) The velocity of the origin of space coordinates (three numbers)	All four functions of position $x^\alpha(\mathscr{P})$
Under what conditions is "$\mathscr{P}$ and $\mathscr{Q}$ are simultaneous" well-defined?	In general; it is a coordinate-free geometric concept	Only after a choice of Lorentz frame has been made; "simultaneity" depends on the frame's velocity	Only after arbitrary choice of time coordinate has been made
Under what conditions is "$\mathscr{P}$ and $\mathscr{Q}$ occur at same point in space" well-defined?	Only after choice of Galilean coordinates has been made	Only after choice of Lorentz coordinates has been made	Only after arbitrary choice of space coordinates has been made
Under what conditions is "$\boldsymbol{u}$ and $\boldsymbol{v}$, at different events, point in same direction" well-defined?	Only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are both spatial vectors ($\langle \boldsymbol{d}t, \boldsymbol{u}\rangle = \langle \boldsymbol{d}t, \boldsymbol{v}\rangle = 0$); or if they lie in the same space slice and are arbitrary vectors; or if there exists a preferred route connecting their locations, along which to compare them by parallel transport	Always	Only if $\boldsymbol{u}$ and $\boldsymbol{v}$ lie at events infinitesimally close together; or if there exists a preferred route (e.g., a unique geodesic) connecting their locations, along which to compare them by parallel transport
Under what conditions is "the invariant distance between $\mathscr{P}$ and $\mathscr{Q}$" well-defined?	Only if $\mathscr{P}$ and $\mathscr{Q}$ lie in the same space slice	Always	Only if $\mathscr{P}$ and $\mathscr{Q}$ are sufficiently close together; or if there exists a unique preferred world line (e.g., a geodesic) linking them, along which to measure the distance

Box 12.3 NEWTONIAN GRAVITY Á LA CARTAN, AND EINSTEINIAN GRAVITY: COMPARISON AND CONTRAST

Property	*Newton-Cartan*	*Einstein*
Idea in brief (formulations of the equivalence principle of very different scope)	Laws of motion of free particles in a local, freely falling, nonrotating frame are identical to Newton's laws of motion as expressed in a gravity-free Galilean frame	Laws of physics in a local, freely falling, nonrotating frame are identical with the laws of physics as formulated in special relativity in a Lorentz frame
Idea even more briefly stated	Point mechanics simple in a local inertial frame	Everything simple in a local inertial frame
Consequence (tested to one part in 10^{11} by Roll-Krotkov-Dicke experiment)	Test particles of diverse composition started with same initial position and same initial velocity follow the same world line ("definition of geodesic")	Test particles of diverse composition started with same initial position and same initial velocity follow the same world line ("definition of geodesic")
Another consequence	In every local region, there exists a local frame ("freely falling frame") in which all geodesics appear straight (all $\Gamma^\alpha{}_{\mu\nu} = 0$)	In every local region there exists a local frame ("freely falling frame") in which all geodesics appear straight (all $\Gamma^\alpha{}_{\mu\nu} = 0$)
Consequence of way light rays travel in real physical world?	Disregarded or evaded. All light rays have same velocity? Speed depend on motion of source? Speed depend on motion of observer? Possible to move fast enough to catch up with a light ray? No satisfactory position on any of these issues	Spacetime always and everywhere has local Lorentz character
Summary of spacetime structure	Stratified into spacelike slices; geometry in each slice Euclidean; each slice characterized by value of universal time (geodesic parameter); displacement of one slice with respect to another not specified; no such thing as a spacetime interval	No stratification. Well-defined interval between every event and every nearby event; spacetime has everywhere local Lorentz character, with one local frame (specific space and time axes) as good as another (other space and time axes); "homogeneous" rather than stratified
This structure expressed in mathematical language	$\Gamma^\alpha{}_{\mu\nu}$'s, yes; spacetime metric $g_{\mu\nu}$, no; $\Gamma^i{}_{00} = \dfrac{\partial\Phi}{\partial x^i}$ $(i = 1, 2, 3)$; all other $\Gamma^\alpha{}_{\mu\nu}$ vanish	$\Gamma^\alpha{}_{\mu\nu}$'s have no independent existence; all derived from $\Gamma^\alpha{}_{\mu\nu} = g^{\alpha\beta}\dfrac{1}{2}\left(\dfrac{\partial g_{\beta\nu}}{\partial x^\mu} + \dfrac{\partial g_{\beta\mu}}{\partial x^\nu} - \dfrac{\partial g_{\mu\nu}}{\partial x^\beta}\right)$ ("metric theory of gravity")

EXERCISES

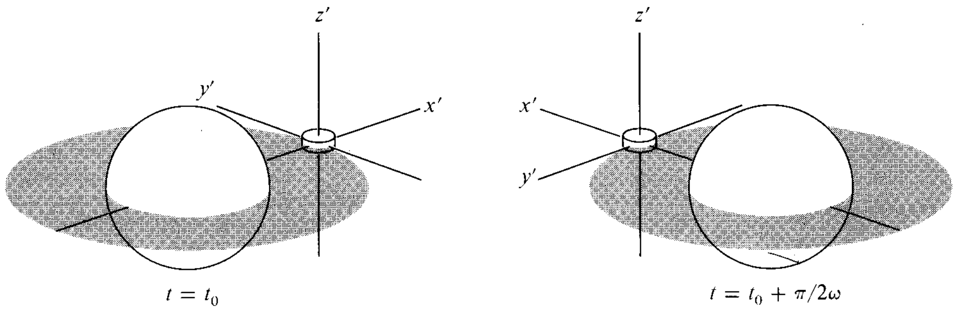


Figure 12.1.
The coordinate system carried by an orbital laboratory as it moves in a circular orbit about the Earth.

Exercise 12.4. CONNECTION COEFFICIENTS FOR ROTATING, ACCELERATING COORDINATES

Beginning with equation (12.4) for the connection coefficients of a Galilean coordinate system $\{x^{\alpha}(\mathscr{P})\}$, derive the connection coefficients (12.14) of the coordinate system $\{x^{\alpha'}(\mathscr{P})\}$ of equations (12.13). From this, verify that (12.15) are necessary and sufficient for $\{x^{\alpha'}(\mathscr{P})\}$ to be Galilean.

Exercise 12.5. EINSTEIN'S ELEVATOR

Use the formalism of this chapter to discuss "Einstein's elevator"—i.e., the equivalence of "gravity" to an acceleration of one's reference frame. Which aspects of "gravity" are equivalent to an acceleration, and which are not?

Exercise 12.6. GEODESIC DEVIATION ABOVE THE EARTH

A manned orbital laboratory is put into a circular orbit about the Earth [radius of orbit $= r_0$, angular velocity $= \omega = (M/r_0{}^3)^{1/2}$—why?]. An astronaut jetisons a bag of garbage and watches it move along its geodesic path. He observes its motion relative to (non-Galilean) space coordinates $\{x^{j'}(\mathscr{P})\}$ which—see Figure 12.1—(1) are Euclidean at each moment of universal time $[(\partial/\partial x^{j'}) \cdot (\partial/\partial x^{k'}) = \delta_{jk}]$, (2) have origin at the laboratory's center, (3) have $\partial/\partial x'$ pointing away from the Earth, (4) have $\partial/\partial x'$ and $\partial/\partial y'$ in the plane of orbit. Use the equation of geodesic deviation to calculate the motion of the garbage bag in this coordinate system. Verify the answer by examining the Keplerian orbits of laboratory and garbage. *Hints:* (1) Calculate $R^{\alpha'}{}_{\beta'\gamma'\delta'}$ in this coordinate system by a trivial transformation of tensorial components. (2) Use equation (12.14) to calculate $\Gamma^{\alpha'}{}_{\beta'\gamma'}$ at the center of the laboratory (i.e., on the fiducial geodesic).

§12.4. GEOMETRIC, COORDINATE-FREE FORMULATION OF NEWTONIAN GRAVITY

To restate Newton's theory of gravity in coordinate-independent, geometric language is the principal goal of this chapter. It has been achieved, thus far, with extensive assistance from a special class of coordinate systems, the Galilean coordinates. To climb out of Galilean coordinates and into completely coordinate-free language is straightforward in principle. One merely passes from index notation to abstract notation.

Example: Restate in coordinate-free language the condition $\Gamma^0{}_{\alpha\beta} = 0$ of Galilean coordinates.

Solution: Write $\Gamma^0{}_{\alpha\beta} = -\langle \boldsymbol{\nabla}_\beta \boldsymbol{d}t, \boldsymbol{e}_\alpha \rangle$; the vanishing of this for all α means $\boldsymbol{\nabla}_\beta \boldsymbol{d}t = 0$ for all β, which in turn means $\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{d}t = 0$ for all $\boldsymbol{u}$. In words: *the gradient of universal time is covariantly constant.*

By this process one can construct a set of coordinate-free statements about Newtonian spacetime (Box 12.4) that are completely equivalent to the standard, non-geometric version of Newton's gravitation theory. From standard Newtonian theory, one can deduce these geometric statements (exercise 12.7); from these geometric statements, regarded as axioms, one can deduce standard Newtonian theory (exercise 12.8).

Coordinate-free, geometric axioms for Newton's theory of gravity

---

EXERCISES

Exercise 12.7. FROM NEWTON TO CARTAN

From the standard axioms of Newtonian theory (last part of Box 12.4) derive the geometric axioms (first part of Box 12.4). *Suggested procedure:* Verify each of the geometric axioms by a calculation in the Galilean coordinate system. Make free use of the calculations and results in §12.1.

Exercise 12.8. FROM CARTAN TO NEWTON

From the geometric axioms of Newtonian theory (first part of Box 12.4) derive the standard axioms (last part of Box 12.4). *Suggested procedure:* (1) Pick three orthonormal, spatial basis vectors ($\boldsymbol{e}_j$ with $\boldsymbol{e}_j \cdot \boldsymbol{e}_k = \delta_{jk}$) at some event $\mathscr{P}_0$. Parallel transport each of them by arbitrary routes to all other events in spacetime.

(2) Use the condition $\mathscr{R}(\boldsymbol{u}, \boldsymbol{n})\boldsymbol{e}_j = 0$ for all $\boldsymbol{u}$ and $\boldsymbol{n}$ [axiom (3)] and an argument like that in §11.5 to conclude: (a) the resultant vector fields $\boldsymbol{e}_i$ are independent of the arbitrary transport routes, (b) $\boldsymbol{\nabla}\boldsymbol{e}_j = 0$ for the resultant fields, and (c) $[\boldsymbol{e}_j, \boldsymbol{e}_k] = 0$.

(3) Pick an arbitrary "time line", which passes through each space slice (slice of constant t) once and only once. Parametrize it by t and select its tangent vector as the basis vector $\boldsymbol{e}_0$ at each event along it. Parallel transport each of these $\boldsymbol{e}_0$'s throughout its respective space slice by arbitrary routes.

(4) From axiom (4) conclude that the resultant field is independent of the transport routes; also show that the above construction process guarantees $\boldsymbol{\nabla}_j \boldsymbol{e}_0 = \boldsymbol{\nabla}_0 \boldsymbol{e}_j = 0$.

(5) Show that $[\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta] = 0$ for all pairs of the four basis-vector fields, and conclude from this that there exists a coordinate system ("Galilean coordinates") in which $\boldsymbol{e}_\alpha = \partial/\partial x^\alpha$ (see §11.5 and exercise 9.9).

(6) Show that in this coordinate system $\boldsymbol{e}_j \cdot \boldsymbol{e}_k = \delta_{jk}$ everywhere (space coordinates are Euclidean), and the only nonzero components of the connection coefficient are $\Gamma^j{}_{00}$; here axioms (6) and (2) will be helpful.

(7) From the self-adjoint property of the Jacobi curvature operator (axiom 7) show that $R^j{}_{0k0} = R^k{}_{0j0}$; show that in terms of the connection coefficients this reads $\Gamma^j{}_{00,k} = \Gamma^k{}_{00,j}$; and from this conclude that there exists a potential Φ such that $\Gamma^j{}_{00} = \Phi_{,j}$.

(8) Show that the geometric field equation (axiom 5) reduces to Poisson's equation $\boldsymbol{\nabla}^2\Phi = 4\pi\rho$.

(9) Show that the geodesic equation for free fall (axiom 8) reduces to the Newtonian equation of motion $d^2x^j/dt^2 + \Phi_{,j} = 0$.

(continued on page 302)

Box 12.4 NEWTONIAN GRAVITY: GEOMETRIC FORMULATION CONTRASTED WITH STANDARD FORMULATION

Geometric Formulation

Newton's theory of gravity and the properties of Newtonian spacetime can be derived from the following axioms. (For derivation see exercise 12.8.)

(1) There exists a function t called "universal time", and a symmetric covariant derivative $\boldsymbol{\nabla}$ (with associated geodesics, parallel transport law, curvature operator, etc.).

(2) The 1-form $\boldsymbol{d}t$ is covariantly constant; i.e.,

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\, \boldsymbol{d}t = 0 \text{ for all } \boldsymbol{u}.$$

[Consequence: if $\boldsymbol{w}$ is a spatial vector field (i.e., $\boldsymbol{w}$ lies everywhere in a surface of constant t; i.e. $\langle \boldsymbol{d}t, \boldsymbol{w}\rangle = 0$ everywhere), then $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}$ is also spatial for every $\boldsymbol{u}$,

$$\langle \boldsymbol{d}t, \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}\rangle = \boldsymbol{\nabla}_{\boldsymbol{u}} \underbrace{\langle \boldsymbol{d}t, \boldsymbol{w}\rangle}_{[0 \text{ always}]} - \langle \underbrace{\boldsymbol{\nabla}_{\boldsymbol{u}}\, \boldsymbol{d}t}_{[0 \text{ always}]}, \boldsymbol{w}\rangle = 0.]$$

(3) Spatial vectors are unchanged by parallel transport around infinitesimal closed curves; i.e.,

$$\mathcal{R}(\boldsymbol{u}, \boldsymbol{n})\boldsymbol{w} = 0 \text{ if } \boldsymbol{w} \text{ is spatial, for every } \boldsymbol{u} \text{ and } \boldsymbol{n}.$$

(4) All vectors are unchanged by parallel transport around infinitesimal, spatial, closed curves; i.e.,

$$\mathcal{R}(\boldsymbol{v}, \boldsymbol{w}) = 0 \text{ for every spatial } \boldsymbol{v} \text{ and } \boldsymbol{w}.$$

(5) The Ricci curvature tensor, $R_{\alpha\beta} \equiv R^{\mu}{}_{\alpha\mu\beta}$, has the form

$$\boldsymbol{Ricci} = 4\pi\rho\, \boldsymbol{d}t \otimes \boldsymbol{d}t,$$

where ρ is the density of mass.

(6) There exists a metric "$\cdot$" defined on spatial vectors only, which is compatible with the covariant derivative in this sense: for any spatial $\boldsymbol{w}$ and $\boldsymbol{v}$, and for any $\boldsymbol{u}$ whatsoever,

$$\boldsymbol{\nabla}_{\boldsymbol{u}}(\boldsymbol{w}\cdot\boldsymbol{v}) = (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w})\cdot\boldsymbol{v} + \boldsymbol{w}\cdot(\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}).$$

[Note: axioms (1), (2), and (3) guarantee that such a spatial metric can exist; see exercise 12.9.]

(7) The Jacobi curvature operator $\mathcal{J}(\boldsymbol{u}, \boldsymbol{n})$, defined for any vectors $\boldsymbol{u}$, $\boldsymbol{n}$, $\boldsymbol{p}$ by

$$\mathcal{J}(\boldsymbol{u}, \boldsymbol{n})\boldsymbol{p} = \frac{1}{2}[\mathcal{R}(\boldsymbol{p}, \boldsymbol{n})\boldsymbol{u} + \mathcal{R}(\boldsymbol{p}, \boldsymbol{u})\boldsymbol{n}],$$

is *"self-adjoint"* when operating on spatial vectors; i.e.,

$$\boldsymbol{v} \cdot [\mathcal{J}(\boldsymbol{u}, \boldsymbol{n})\boldsymbol{w}] = \boldsymbol{w} \cdot [\mathcal{J}(\boldsymbol{u}, \boldsymbol{n})\boldsymbol{v}] \text{ for all spatial } \boldsymbol{v}, \boldsymbol{w}; \text{ and for any } \boldsymbol{u}, \boldsymbol{n}.$$

(8) "Ideal rods" measure the lengths that are calculated with the spatial metric; "ideal clocks" measure universal time t (or some multiple thereof); and "freely falling particles" move along geodesics of $\boldsymbol{\nabla}$. [Note: this can be regarded as a definition of "ideal rods," "ideal clocks," and "freely falling particles." A more complete theory (e.g., general relativity; see §16.4) would predict in advance whether a given physical rod or clock is ideal, and whether a given real particle is freely falling.]

Note: For an alternative but equivalent set of axioms, see pp. 106–107 of Trautman (1965).

Standard Formulation

The following standard axioms are equivalent to the above.

(1) There exist a universal time t, a set of Cartesian space coordinates x^j (called "Galilean coordinates"), and a Newtonian gravitational potential Φ.
(2) The density of mass ρ generates the Newtonian potential by Poisson's equation,

$$\boldsymbol{\nabla}^2\Phi \equiv \frac{\partial^2\Phi}{\partial x^j\,\partial x^j} = 4\pi\rho.$$

(3) The equation of motion for a freely falling particle is

$$\frac{d^2x^j}{dt^2} + \frac{\partial\Phi}{\partial x^j} = 0.$$

(4) "Ideal rods" measure the Galilean coordinate lengths; "ideal clocks" measure universal time.

Exercise 12.9. SPATIAL METRIC ALLOWED BY OTHER AXIOMS

Show that the geometric axioms (1), (2), and (3) of Box 12.4 permit one to introduce a spatial metric satisfying axiom (6). *Hint:* Pick an arbitrary spatial basis $\{\boldsymbol{e}_j\}$ at some event. *Define* it to be orthonormal, $\boldsymbol{e}_j \cdot \boldsymbol{e}_k \equiv \delta_{jk}$. Extend this basis through all spacetime by the method used in (1) of exercise 12.8. Define $\boldsymbol{e}_j \cdot \boldsymbol{e}_k \equiv \delta_{jk}$ everywhere in spacetime for this basis. Then prove that the resulting metric satisfies the compatibility condition of axiom (6).

Exercise 12.10. SPACETIME METRIC FORBIDDEN BY OTHER AXIOMS

Show that in Newtonian spacetime it is impossible to construct a nondegenerate spacetime metric $\boldsymbol{g}$, defined on *all* vectors, that is compatible with the covariant derivative in the sense that

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{g}(\boldsymbol{n}, \boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{n}, \boldsymbol{p}) + \boldsymbol{g}(\boldsymbol{n}, \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{p}). \tag{12.20}$$

Note: to prove this requires mastery of the material in Chapter 8 or 13; so study either 8 or 13 before tackling it. *Hint:* Assume that such a $\boldsymbol{g}$ exists. Show, by the methods of exercise 12.8, that in a Galilean coordinate system the spatial components g_{jk} are independent of position in spacetime. Then use this and the form of $R^{\alpha}{}_{\beta\gamma\delta}$ in Galilean coordinates to prove R_{i0k0} and $-R_{0ik0}$ are not identical, a result that conflicts with the symmetries of the Riemann tensor [eq. (8.45)] in a manifold with compatible metric and covariant derivative.

§12.5. THE GEOMETRIC VIEW OF PHYSICS: A CRITIQUE

An important digression is in order.

The principle of general covariance has no forcible content

"Every physical quantity must be describable by a (coordinate-free) geometric object, and the laws of physics must all be expressible as geometric relationships between these geometric objects." This view of physics, sometimes known as the *"principle of general covariance,"* pervades twentieth-century thinking. But does it have any forcible content? No, not at all, according to one viewpoint that dates back to Kretschmann (1917). Any physical theory originally written in a special coordinate system can be recast in geometric, coordinate-free language. Newtonian theory is a good example, with its equivalent geometric and standard formulations (Box 12.4). Hence, as a sieve for separating viable theories from nonviable theories, the principle of general covariance is useless.

Twentieth-century viewpoint judges a theory by simplicity of its geometric formulation

Einstein's theory of gravity is simple; Newton's is complex

But another viewpoint is cogent. It constructs a powerful sieve in the form of a slightly altered and slightly more nebulous principle: "Nature likes theories that are simple when stated in coordinate-free, geometric language."* According to this principle, Nature must love general relativity, and it must hate Newtonian theory. Of all theories ever conceived by physicists, general relativity has the simplest, most elegant geometric foundation [three axioms: (1) there is a metric; (2) the metric is governed by the Einstein field equation $\boldsymbol{G} = 8\pi\boldsymbol{T}$; (3) all special relativistic laws of physics are valid in local Lorentz frames of metric]. By contrast, what diabolically

*Admittedly, this principle is anthropomorphic: twentieth-century physicists like such theories and even find them effective in correlating observational data. Therefore, Nature must like them too!

clever physicist would ever foist on man a theory with such a complicated geometric foundation as Newtonian theory?

Of course, from the nineteenth-century viewpoint, the roles are reversed. It judged simplicity of theories by examining their coordinate formulations. In Galilean coordinates, Newtonian theory is beautifully simple. Expressed as differential equations for the metric coefficients in a specific coordinate system, Einstein's field equations (ten of them now!) are horrendously complex.

The geometric, twentieth-century view prevails because it accords best with experimental data (see Chapters 38–40). In Chapter 17 it will be applied ruthlessly to make Einstein's field equation seem compelling.

2

CHAPTER 13

RIEMANNIAN GEOMETRY: METRIC AS FOUNDATION OF ALL

Philosophy is written in this great book (by which I mean the universe) which stands always open to our view, but it cannot be understood unless one first learns how to comprehend the language and interpret the symbols in which it is written, and its symbols are triangles, circles, and other geometric figures, without which it is not humanly possible to comprehend even one word of it; without these one wanders in a dark labyrinth.

GALILEO GALILEI (1623)

§13.1. NEW FEATURES IMPOSED ON GEOMETRY BY LOCAL VALIDITY OF SPECIAL RELATIVITY

This chapter is entirely Track 2. Chapters 9–11 are necessary preparation for it. It will be needed as preparation for
(1) the second half of Chapter 14 (calculation of curvature), and
(2) the details, but not the message, of Chapter 15 (Bianchi identities).
§13.6 (proper reference frame) will be useful throughout the applications of gravitation theory (Chapters 18–40).

Constraints imposed on spacetime by special relativity

Freely falling particles (geodesics) define and probe the structure of spacetime. This spacetime is curved. Gravitation is a manifestation of its curvature. So far, so good, in terms of Newton's theory of gravity as translated into geometric language by Cartan. What is absolutely unacceptable, however, is the further consequence of the Cartan-Newton viewpoint (Chapter 12): stratification of spacetime into slidable slices, with no meaning for the spacetime separation between an event in one slice and an event in another.

Of all the foundations of physics, none is more firmly established than special relativity; and of all the lessons of special relativity none stand out with greater force than these. (1) Spacetime, far from being stratified, is homogeneous and isotropic throughout any region small enough ("local region") that gravitational tide-producing effects ("spacetime curvatures") are negligible. (2) No local experiment whatsoever can distinguish one local inertial frame from another. (3) The speed of light is the same in every local inertial frame. (4) It is not possible to give frame-independent meaning to the separation in time ("no Newtonian stratification"). (5) Between every event and every nearby event there exists a frame-independent, coordinate-independent spacetime interval ("Riemannian geometry"). (6) Spacetime is always and everywhere locally Lorentz in character ("local Lorentz character of this Riemannian geometry").

Metric: the instrument which imposes the constraints

What mathematics gives all these physical properties? A metric; a metric that is locally Lorentz (§§13.2 and 13.6). All else then follows. In particular, the metric destroys the stratified structure of Newtonian spacetime, as well as its gravitational potential and universal time coordinate. But not destroyed are the deepest features of Newtonian gravity: (1) the equivalence principle (as embodied in geodesic description of free-fall motion, §§13.3 and 13.4); and (2) spacetime curvature (as measured by tidal effects, §13.5).

The skyscraper of vectors, forms, tensors (Chapter 9), geodesics, parallel transport, covariant derivative (Chapter 10), and curvature (Chapter 11) has rested on crumbling foundations—Newtonian physics and a geodesic law based on Newtonian physics. But with metric now on the scene, the whole skyscraper can be transferred to new foundations without a crack appearing. Only one change is necessary: the geodesic law must be selected in a new, relativistic way; a way based on metric (§§13.3 and 13.4). Resting on metric foundations, spacetime curvature acquires additional and stronger properties (the skyscraper is redecorated and extended), which are studied in §13.5 and in Chapters 14 and 15, and which lead almost inexorably to Einstein's field equation.

§13.2. METRIC

A spacetime metric; a curved spacetime metric; a locally Lorentz, curved spacetime metric. This is the foundation of spacetime geometry in the real, physical world. Therefore take a moment to recall what "metric" is in three contrasting languages.

Metric described in three languages

In the language of elementary geometry, "metric" is a table giving the interval between every event and every other event (Box 13.1 and Figure 13.1). In the language of coordinates, "metric" is a set of ten functions of position, $g_{\mu\nu}(x^\alpha)$, such that the expression

$$\Delta s^2 = -\Delta\tau^2 = g_{\mu\nu}(x^\alpha)\,\Delta x^\mu\,\Delta x^\nu \tag{13.1}$$

gives the interval between any event x^α and any nearby event $x^\alpha + \Delta x^\alpha$. In the language of abstract differential geometry, metric is a bilinear machine, $\boldsymbol{g} \equiv (\ldots\cdot\ldots)$, to produce a number ["scalar product $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) \equiv (\boldsymbol{u}\cdot\boldsymbol{v})$"] out of two tangent vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$.

The link between the abstract, machine viewpoint and the concrete coordinate viewpoint is readily exhibited. Let the tangent vector

$$\boldsymbol{\xi} \equiv \Delta x^\alpha \boldsymbol{e}_\alpha = \Delta x^\alpha(\partial/\partial x^\alpha)$$

represent the displacement between two neighboring events. The abstract viewpoint gives

$$\Delta s^2 \equiv \boldsymbol{\xi}\cdot\boldsymbol{\xi} \equiv \boldsymbol{g}(\Delta x^\mu \boldsymbol{e}_\mu, \Delta x^\nu \boldsymbol{e}_\nu) = \Delta x^\mu\,\Delta x^\nu \boldsymbol{g}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)$$

(*continued on page 310*)

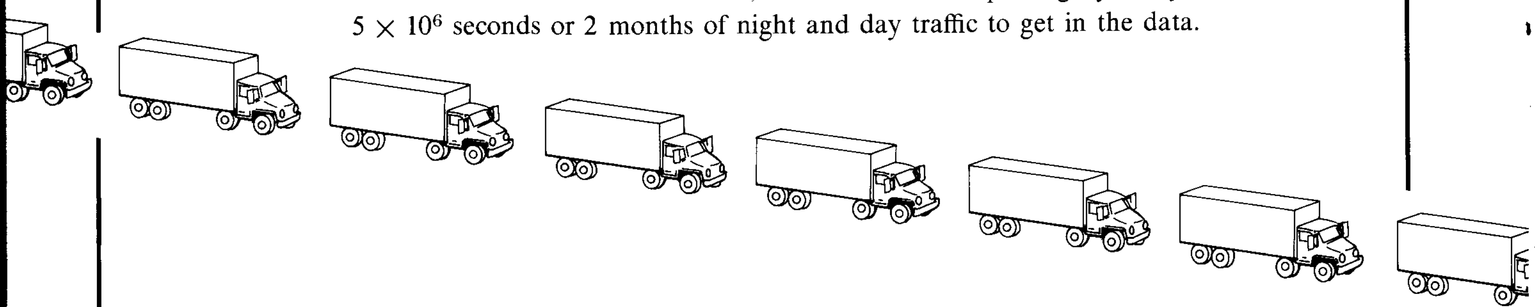
Box 13.1 METRIC DISTILLED FROM DISTANCES

Raw Data on Distances

Let the shape of the earth be described as in Figure 13.1, by giving distances between some of the principal identifiable points: buoys, ships, icebergs, lighthouses, peaks, and flags: points to a total of $n = 2 \times 10^7$. The total number of distances to be given is $n(n-1)/2 = 2 \times 10^{14}$. With 200 distances per page of printout, this means

First point	Second point	Distance (Nautical miles)	First point	Seco pr
9,316,434	14,117,103	1410.316	9,316,434	
9,316,434	14,117,104	1812.717	9,316,434	
9,316,434	14,117,105	1629.291	9.3	
[illegible]	[illegible]			

10^{12} pages weighing 6 g each, or 6×10^6 metric tons of data. With 6 tons per truck this means 10^6 truckloads of data; or with one truck passing by every 5 seconds, 5×10^6 seconds or 2 months of night and day traffic to get in the data.

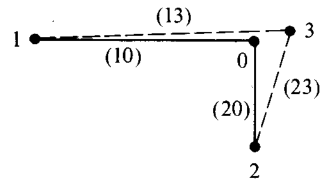


First Distillation: Distances to Nearby Points Only

Get distances between faraway points by adding distances covered on the elementary short legs of the trip. Boil down the table of distances to give only the distance between each point and the hundred nearest points. Now have $100\,n = 2 \times 10^9$ distances, or $2 \times 10^9/200 = 10^7$ pages of data, or 60 tons of records, or 10 truckloads.

Second Distillation: Distances Between Nearby Points in Terms of Metric

Idealize the surface of the earth as smooth. Then in any sufficiently limited region the geometry is Euclidean. This circumstance has a happy consequence. It is enough to know a few distances between the nearby points to be able to determine all the



distances between the nearby points. Locate point 2 so that (102) is a right triangle; thus, $(12)^2 = (10)^2 + (20)^2$. Consider a point 3 close to 0. Define

$$x(3) = (13) - (10)$$

and

$$y(3) = (23) - (20).$$

Then the distance (03) does not have to be supplied independently; it can be calculated from the formula*

$$(03)^2 = [x(3)]^2 + [y(3)]^2.$$

Similarly for a point 4 and its distance (04) from the local origin 0. Similarly for the distance (mn) between any two points m and n that are close to 0:

$$(mn)^2 = [x(m) - x(n)]^2 + [y(m) - y(n)]^2.$$

Thus it is only needful to have the distance $(1m)$ (from point 1) and $(2m)$ (from point 2) for each point m close to 0 ($m = 3, 4, \ldots, N + 2$) to be able to work out

*If the distance (03) is given arbitrarily, the resulting four-vertex figure will burst out of the plane. Regarded as a tetrahedron in a three-dimensional Euclidean space, it has a volume given by the formula of Niccolo Fontana Tartaglia (1500–1557), generalized today (Blumenthal 1953) to

$$\begin{pmatrix}\text{volume of}\\ n\text{-dimensional}\\ \text{simplex}\\ \text{spanned by}\\ (n+1)\text{ points}\end{pmatrix} = \left(\frac{(-1)^{n+1}}{2^n}\right)^{1/2}\frac{1}{n!}\begin{vmatrix} 0 & 1 & 1 & 1 & \ldots & 1\\ 1 & 0 & (01)^2 & (02)^2 & \ldots & (0n)^2\\ 1 & (10)^2 & 0 & (12)^2 & \ldots & (1n)^2\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & (n0)^2 & (n1)^2 & (n2)^2 & \ldots & 0\end{vmatrix}^{1/2},$$

which reduces for three points to the standard textbook formula of Hero of Alexandria (A.D. 62 to A.D. 150).

$$\text{area} = \{s[s - (01)][s - (02)][s - (12)]\}^{1/2},$$
$$2s = (01) + (02) + (12),$$

for the area of a triangle. Conversely, if the four points are to remain in two-dimensional Euclidean space, the tetrahedron must collapse to zero volume. This requirement supplies one condition on the one distance (03). It simplifies the discussion of this condition to take (03) small and (102) to be a right triangle, as above. However, the general principle is independent of such approximations, and follows directly from the extended Hero-Tartaglia formula. It is enough in a locally Euclidean or Lorentz space of n dimensions to have laid down $(n + 1)$ fiducial points $0, 1, 2, \ldots, n$, and to know the distance of every other point $j, k, \ldots$ from these fiducial points, in order to be able to calculate the distance of these points $j, k, \ldots$ from one another ("distances between nearby points in terms of coordinates"; metric as distillation of distance data).

Box 13.1 *(continued)*

its distance from every point n close to 0. The prescription to determine the $N(N-1)/2$ distances between these N nearby points can be reexpressed to advantage in these words: (1) each point has two coordinates, x and y; and (2) the distance is given in terms of these coordinates by the standard Euclidean metric; thus

$$(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2.$$

Having gone this far on the basis of "distance geometry" (for more on which, see Robb 1914 and 1936), one can generalize from a small region (Euclidean) to a large region (not Euclidean). Introduce any arbitrary smooth pair of everywhere-independent curvilinear coordinates x^k, and express distance, not only in the immediate neighborhood of the point 0, but also in the immediate neighborhood of every point of the surface (except places where one has to go to another coordinate patch; at least two patches needed for 2-sphere) in terms of the formula

$$ds^2 = g_{jk}\, dx^j\, dx^k.$$

Thus out of the table of distances between nearby points one has distilled now five numbers per point (two coordinates, x^1, x^2, and three metric coefficients, g_{11}, $g_{12} = g_{21}$, and g_{22}), down by a factor of $100/5 = 20$ from what one had before (now 3 tons of data, or half a truckload).

Third Distillation: Metric Coefficients Expressed as Analytical Functions of Coordinates

Instead of giving the three metric coefficients at each of the 2×10^7 points of the surface, give them as functions of the two coordinates x^1, x^2, in terms of a power series or an expansion in spherical harmonics or otherwise with some modest number, say 100, of adjustable coefficients. Then the information about the geometry itself (as distinct from the coordinates of the 2×10^7 points located on that geometry) is caught up in these three hundred coefficients, a single page of printout. Goodbye to any truck! In brief, metric provides a shorthand way of giving the distance between every point and every other point—but its role, its justification and its meaning lies in these distances and only in these many distances.

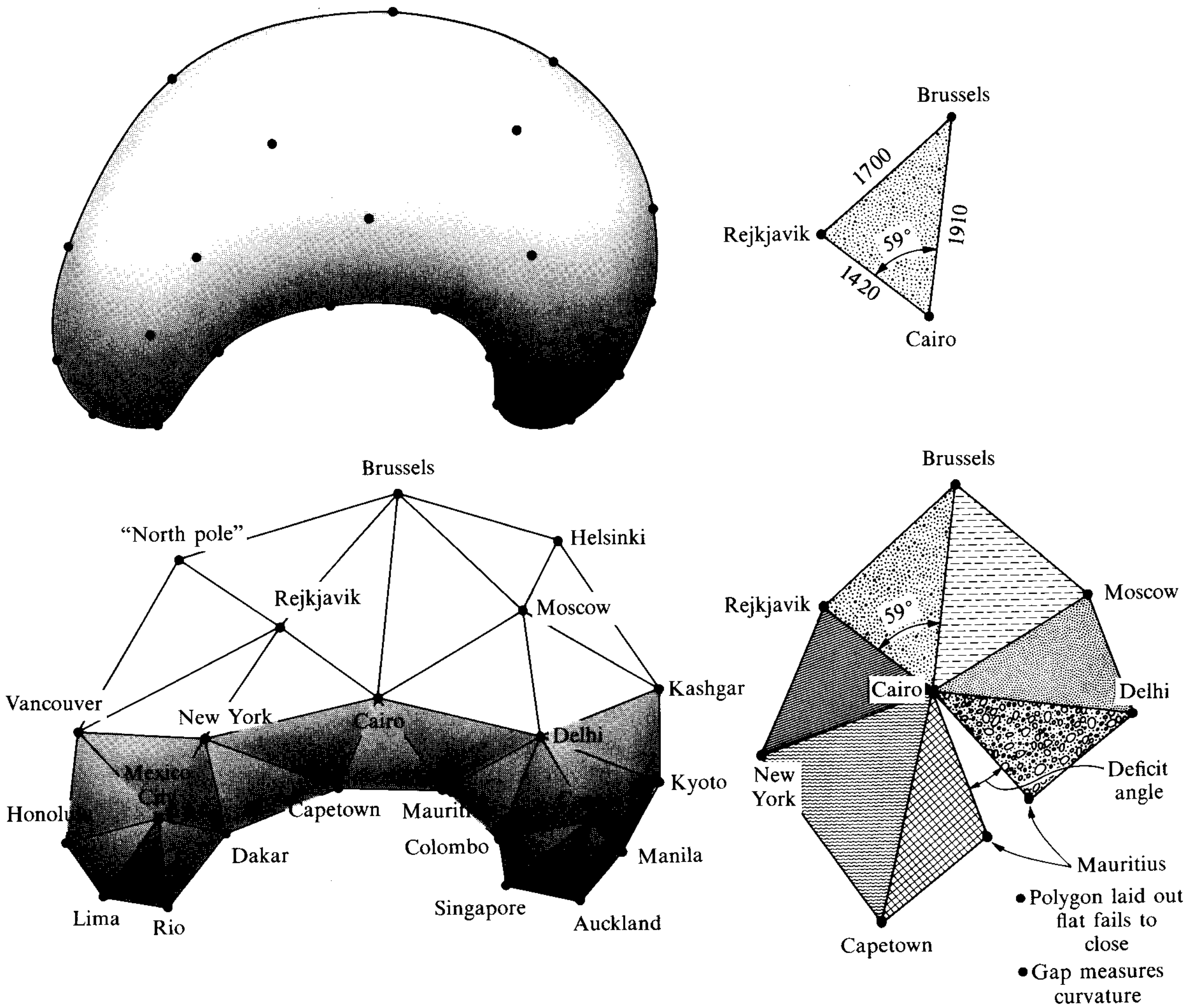


Figure 13.1.
Distances determine geometry. Upper left: Sufficiently great tidal forces, applied to the earth with tailored timing, have deformed it to the shape of a tear drop. Lower left: This tear drop is approximated by a polyhedron built out of triangles ("skeleton geometry"). The approximation can be made arbitrarily good by making the number of triangles sufficiently great and the size of each sufficiently small. Upper right: The geometry in each triangle is Euclidean: giving the three edge lengths fixes all the features of the figure, including the indicated angle. Lower right: The triangles that belong to a given vertex, laid out on a flat surface, fail to meet. The deficit angle measures the amount of curvature concentrated at that vertex on the tear-drop earth. The sum of these deficit angles for all vertices of the tear drop equals 4π. This "Gauss-Bonnet theorem" is valid for any figure with the topology of the 2-sphere; for the simplest figure of all, a tetrahedron, four vertices with a deficit angle at each of 180° are needed—3 triangles $\times$ 60° per triangle available = 180° deficit. In brief, the shape of the tear drop, in the given skeleton-geometry approximation, is determined by its 50 visible edge lengths plus, say, 32 more edge lengths hidden behind the figure, or a total of 82 edge lengths, and by nothing more ("distances determine geometry"). "Metric" tells the distance between every point and every nearby point. If volcanic action raises Rejkjavik, the distances between that Icelandic capital and nearby points increase accordingly; distances again reveal shape. Conversely, that there is not a great bump on the earth in the vicinity of Iceland, and that the earth does not now have a tear-drop shape, can be unambiguously established by analyzing the pattern of distances from point to point in a sufficiently well-distributed network of points, with no call for any observations other than measurements of distance.

2

for the interval between those events; comparison with the coordinate viewpoint [equation (13.1)] reveals

Covariant components of metric

$$g_{\mu\nu} = \boldsymbol{g}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu \tag{13.2}$$

(standard equation for calculating components of a tensor).

"Line element" compared with "metric as bilinear machine"

Just as modern differential geometry replaces the old style "differential" df by the "differential form" $\boldsymbol{d}f$ (Box 2.3, page 63), so it also replaces the old-style "line element"

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = (\text{"interval between } x^\alpha \text{ and } x^\alpha + dx^\alpha\text{"}) \tag{13.3}$$

by the bilinear machine ("metric tensor")

$$\boldsymbol{g} \equiv \boldsymbol{ds}^2 \equiv g_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu. \tag{13.4}$$

The output $\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\xi})$ of this machine, for given displacement-vector input, is identical to the old-style interval. Hence, $\boldsymbol{ds}^2 = g_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$ represents the interval of an unspecified displacement; and the act of inserting $\boldsymbol{\xi}$ into the slots of $\boldsymbol{ds}^2$ is the act of making explicit the interval $\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\xi}) = g_{\mu\nu}\, \Delta x^\mu\, \Delta x^\nu$ of an explicit displacement.

Metric produces a correspondence between 1-forms and tangent vectors

In curved spacetime with metric, just as in flat spacetime with metric (§2.5), a particular 1-form $\tilde{\boldsymbol{u}}$ corresponds to any given tangent vector $\boldsymbol{u}$:

$$\tilde{\boldsymbol{u}} \text{ is defined by } \text{"}\langle \tilde{\boldsymbol{u}}, \boldsymbol{v}\rangle \equiv \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) \text{ for all } \boldsymbol{v}\text{"} \tag{13.5}$$

("representation of the same physical quantity in the two alternative versions of vector and 1-form"; "corresponding representations" as $\binom{1}{0}$-tensor and as $\binom{0}{1}$-tensor). Example: the 1-form $\tilde{\boldsymbol{u}}$ corresponding to a basis vector $\boldsymbol{u} = \boldsymbol{e}_\alpha$ has components

$$u_\beta \underset{\substack{\uparrow \\ \left[\begin{array}{l}\text{standard way}\\ \text{to compute } u_\beta\end{array}\right]}}{=} \langle \tilde{\boldsymbol{u}}, \boldsymbol{e}_\beta\rangle \underset{\substack{\uparrow \\ \text{[definition (13.5)]}}}{\equiv} \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{e}_\beta) \underset{\substack{\uparrow \\ [\text{by } \boldsymbol{u} = \boldsymbol{e}_\alpha]}}{=} \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) \underset{\substack{\uparrow \\ \text{[equation (13.2)]}}}{=} g_{\alpha\beta};$$

thus

$$g_{\alpha\beta}\boldsymbol{\omega}^\beta \text{ is the 1-form } \tilde{\boldsymbol{e}}_\alpha \text{ corresponding to } \boldsymbol{e}_\alpha. \tag{13.6}$$

Also as in flat spacetime (§3.2), a tensor can accept either a vector or a 1-form into any given slot

$$\boldsymbol{S}(\tilde{\boldsymbol{u}}, \boldsymbol{\sigma}, \boldsymbol{v}) \equiv \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{\sigma}, \boldsymbol{v}). \tag{13.7}$$

Lowering indices

Equivalently, in component language, the indices of a tensor can be lowered with the covariant components of the metric

$$S_\alpha{}^\beta{}_\gamma \underset{\substack{\uparrow \\ [\text{definition of } S_\alpha{}^\beta{}_\gamma]}}{=} \boldsymbol{S}(\boldsymbol{e}_\alpha, \boldsymbol{\omega}^\beta, \boldsymbol{e}_\gamma) = \boldsymbol{S}(\tilde{\boldsymbol{e}}_\alpha, \boldsymbol{\omega}^\beta, \boldsymbol{e}_\gamma) \underset{\substack{\uparrow \\ \text{[by equation (13.6)]}}}{=} \boldsymbol{S}(g_{\alpha\mu}\boldsymbol{\omega}^\mu, \boldsymbol{\omega}^\beta, \boldsymbol{e}_\gamma) = g_{\alpha\mu} S^{\mu\beta}{}_\gamma. \tag{13.8}$$

The basis vectors $\{\boldsymbol{e}_\alpha\}$ can be chosen arbitrarily at each event. Therefore the corresponding components $g_{\alpha\beta}$ of the metric are quite arbitrary (though symmetric: $g_{\alpha\beta} = g_{\beta\alpha}$). But the mixed components $g^\alpha{}_\beta$ are not arbitrary. In particular, equations (13.5) and (13.7) imply

$$\boldsymbol{g}(\tilde{\boldsymbol{u}}, \boldsymbol{v}) \equiv \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) \equiv \langle \tilde{\boldsymbol{u}}, \boldsymbol{v} \rangle. \tag{13.9}$$

Mixed and contravariant components of metric

Therefore one concludes that the metric tensor in mixed representation is identical with the unit matrix:

$$g^\alpha{}_\beta \equiv \boldsymbol{g}(\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta) \equiv \langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta. \tag{13.10}$$

This feature of the metric in turn fixes the contravariant components of the metric:

$$g^{\alpha\mu} g_{\mu\beta} \underset{\text{["lowering an index" of } g^{\alpha\mu}]}{=} g^\alpha{}_\beta = \delta^\alpha{}_\beta; \tag{13.11}$$

i.e.,

$$\|g^{\alpha\beta}\| \text{ is the matrix inverse of } \|g_{\alpha\beta}\|. \tag{13.12}$$

Raising indices

This reciprocity enables one to undo the lowering of tensor indices (i.e., raise indices) with $g^{\alpha\beta}$:

$$S^{\mu\beta}{}_\gamma = \delta^\mu{}_\alpha S^{\alpha\beta}{}_\gamma = g^{\mu\nu} g_{\nu\alpha} S^{\alpha\beta}{}_\gamma = g^{\mu\nu} S_\nu{}^\beta{}_\gamma. \tag{13.13}$$

The last two paragraphs may be summarized in brief:

(1) $g^\alpha{}_\beta = \delta^\alpha{}_\beta$;
(2) $\|g^{\alpha\beta}\| = \|g_{\alpha\beta}\|^{-1}$;
(3) tensor indices are lowered with $g_{\alpha\beta}$;
(4) tensor indices are raised with $g^{\alpha\beta}$.

Metric must be locally Lorentz

In this formalism of metric and index shuffling, a big question demands attention: how can one tell whether the metric is locally Lorentz rather than locally Euclidean or locally something else? Of course, one criterion (necessary; not sufficient!) is dimensionality—a locally Lorentz spacetime must have four dimensions. (Recall the method of §1.2 to determine dimensionality.) Confine attention, then, to four-dimensional manifolds. What else must one demand? One must demand that at every event $\mathscr{P}$ there exist an orthonormal frame (orthonormal set of basis vectors $\{\boldsymbol{e}_{\hat{\alpha}}\}$) in which the components of the metric have their flat-spacetime form

$$g_{\hat{\alpha}\hat{\beta}} \equiv \boldsymbol{e}_{\hat{\alpha}} \cdot \boldsymbol{e}_{\hat{\beta}} = \eta_{\alpha\beta} \equiv \text{diagonal}\,(-1, 1, 1, 1). \tag{13.14}$$

To test for this is straightforward (exercise 13.1). (1) Search for a timelike vector $\boldsymbol{u}$ ($\boldsymbol{u} \cdot \boldsymbol{u} < 0$). If none exist, spacetime is not locally Lorentz. If one is found, then (2) examine all non-zero vectors $\boldsymbol{v}$ perpendicular to $\boldsymbol{u}$. If they are all spacelike ($\boldsymbol{v} \cdot \boldsymbol{v} > 0$), then spacetime is locally Lorentz. Otherwise it is not.

EXERCISES

Exercise 13.1. TEST WHETHER SPACETIME IS LOCAL LORENTZ

Prove that the above two-step procedure for testing whether spacetime is locally Lorentz is valid: i.e., prove that if the procedure says "yes," then there exists an orthonormal basis with $g_{\hat{\alpha}\hat{\beta}} = \eta_{\alpha\beta}$ at the event in question; if it says "no," then no such basis exists.

Exercise 13.2. PRACTICE WITH METRIC

A four-dimensional manifold with coordinates v, r, θ, ϕ has line element (old-style notation)

$$ds^2 = -(1 - 2M/r)\, dv^2 + 2\, dv\, dr + r^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

corresponding to metric (new-style notation)

$$\boldsymbol{ds^2} = -(1 - 2M/r)\, \boldsymbol{dv} \otimes \boldsymbol{dv} + \boldsymbol{dv} \otimes \boldsymbol{dr} + \boldsymbol{dr} \otimes \boldsymbol{dv} + r^2(\boldsymbol{d\theta} \otimes \boldsymbol{d\theta} + \sin^2\theta\, \boldsymbol{d\phi} \otimes \boldsymbol{d\phi}),$$

where M is a constant.

(a) Find the *"covariant"* components $g_{\alpha\beta}$ and *"contravariant"* components $g^{\alpha\beta}$ of the metric in this coordinate system. [Answer: $g_{vv} = -(1 - 2M/r)$, $g_{vr} = g_{rv} = 1$, $g_{\theta\theta} = r^2$, $g_{\phi\phi} = r^2 \sin^2\theta$; all other $g_{\alpha\beta}$ vanish; $g^{vr} = g^{rv} = 1$, $g^{rr} = (1 - 2M/r)$, $g^{\theta\theta} = r^{-2}$, $g^{\phi\phi} = r^{-2}\sin^{-2}\theta$, all other $g^{\alpha\beta}$ vanish.]

(b) Define a scalar field t by

$$t \equiv v - r - 2M \ln[(r/2M) - 1].$$

What are the covariant and contravariant components (u_α and u^α) of the 1-form $\tilde{\boldsymbol{u}} \equiv \boldsymbol{d}t$? What is the squared length $\boldsymbol{u}^2 \equiv \boldsymbol{u} \cdot \boldsymbol{u}$, of the corresponding vector? Show that $\boldsymbol{u}$ is timelike in the region $r > 2M$. [Answer: $u_v = 1$, $u_r = -1/(1 - 2M/r)$, $u_\theta = u_\phi = 0$; $u^v = -1/(1 - 2M/r)$, $u^r = 0$, $u^\theta = u^\phi = 0$; $\boldsymbol{u}^2 = -1/(1 - 2M/r)$.]

(c) Find the most general non-zero vector $\boldsymbol{w}$ orthogonal to $\boldsymbol{u}$ in the region $r > 2M$, and show that it is spacelike. Thereby conclude that spacetime is locally Lorentz in the region $r > 2M$. [Answer: Since $\boldsymbol{w} \cdot \boldsymbol{u} = w_\alpha u^\alpha = -w_v/(1 - 2M/r)$, w_v must vanish, but w_r, w_θ, w_ϕ are arbitrary, and $\boldsymbol{w}^2 = (1 - 2M/r) w_r{}^2 + r^{-2} w_\theta{}^2 + r^{-2}\sin^{-2}\theta\, w_\phi{}^2 > 0$.]

(d) Let t, r, θ, ϕ be new coordinates for spacetime. Find the line element in this coordinate system. [Answer: This is the "Schwarzschild" line element

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{1 - 2M/r} + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2.]$$

(e) Find an orthonormal basis, for which $g_{\hat{\alpha}\hat{\beta}} = \eta_{\alpha\beta}$ in the region $r > 2M$. [Answer: $\boldsymbol{e}_{\hat{0}} \equiv (1 - 2M/r)^{-1/2}\, \partial/\partial t$, $\boldsymbol{e}_{\hat{r}} \equiv (1 - 2M/r)^{1/2}\, \partial/\partial r$, $\boldsymbol{e}_{\hat{\theta}} \equiv r^{-1}\, \partial/\partial\theta$, $\boldsymbol{e}_{\hat{\phi}} = (r \sin\theta)^{-1}\, \partial/\partial\phi$.]

§13.3. CONCORD BETWEEN GEODESICS OF CURVED SPACETIME GEOMETRY AND STRAIGHT LINES OF LOCAL LORENTZ GEOMETRY

More could be said about the mathematical machinery and physical implications of "metric," but an issue of greater urgency presses for attention. What has metric (or spacetime interval) to do with geodesic (or world line of test particle)? Answer:

Two mathematical objects ("straight line in a local Lorentz frame" and "geodesic of the over-all global curved spacetime geometry") equal to the same physical object ("world line of test particle") must be equal to each other ("condition of consistency"). As a first method to spell out this consistency requirement, examine the two mathematical representations of the world line of a test particle in the neighborhood of a given event $\mathscr{P}_0$. The local-Lorentz representation says:

Local-Lorentz description of straight lines

"Pick a local Lorentz frame at $\mathscr{P}_0$. [As spelled out in exercise 13.3, such a local Lorentz frame is the closest thing there is to a global Lorentz frame at $\mathscr{P}_0$; i.e., it is a coordinate system in which

$$g_{\alpha\beta}(\mathscr{P}_0) = \eta_{\alpha\beta} \text{ (flat-spacetime metric)}, \tag{13.15a}$$

$$g_{\alpha\beta,\gamma}(\mathscr{P}_0) = 0, \tag{13.15b}$$

$$g_{\alpha\beta,\gamma\delta}(\mathscr{P}_0) \neq 0 \text{ except in special cases, such as flat space.]} \tag{13.15c}$$

The world line in that frame has zero acceleration,

$$d^2x^\alpha/d\tau^2 = 0 \text{ at } \mathscr{P}_0 \text{ ("straight-line equation")}, \tag{13.16}$$

where τ is proper time as measured by the particle's clock."

The geodesic representation says

Geodesic description of straight lines

"In the local Lorentz frame, as in any coordinate frame, the world line satisfies the geodesic equation

$$d^2x^\alpha/d\tau^2 + \Gamma^\alpha{}_{\beta\gamma}(dx^\beta/d\tau)(dx^\gamma/d\tau) = 0 \tag{13.17}$$

(τ is an affine parameter because it is time as measured by the test particle's clock)."

Condition of consistency: $\Gamma^\alpha{}_{\beta\gamma} = 0$ in local Lorentz frame

Consistency of the two representations for any and every choice of test particle (any and every choice of $dx^\alpha/d\tau$ at $\mathscr{P}_0$) demands

$$\Gamma^\alpha{}_{\beta\gamma}(\mathscr{P}_0) = 0 \text{ in any local Lorentz frame [coordinate system satisfying equations (13.15) at } \mathscr{P}_0\text{]}; \tag{13.18}$$

i.e., it demands that *every local Lorentz frame is a local inertial frame*. (On local inertial frames see §11.6.) In such a frame, all local effects of "gravitation" disappear. That is the physical shorthand for (13.18).

One does not have to speak in the language of a specific coordinate system when one demands identity between the geodesic (derived from the $\Gamma^\alpha{}_{\beta\gamma}$) and the straight line of the local Lorentz geometry ($g_{\mu\nu}$). The local Lorentz specialization of coordinates may be the most immediate way to see the physics ("no local effects of gravitation"), but it is not the right way to formulate the basic mathematical requirement in its full generality and power. The right way is to demand

Consistency reformulated: $\boldsymbol{\nabla g} = 0$.

$$\boldsymbol{\nabla g} = 0 \text{ ("compatibility of } \boldsymbol{g} \text{ and } \boldsymbol{\nabla}\text{")}. \tag{13.19}$$

Stated in the language of an arbitrary coordinate system, this requirement reads

$$g_{\alpha\beta;\gamma} \equiv \frac{\partial g_{\alpha\beta}}{\partial x^\gamma} - \Gamma^\mu{}_{\alpha\gamma} g_{\mu\beta} - \Gamma^\mu{}_{\beta\gamma} g_{\alpha\mu} = 0. \tag{13.19'}$$

That this covariant requirement is fulfilled in every coordinate system follows from its validity in one coordinate system: a local Lorentz frame. (The first term in this equation, and the last two terms, are separately required to vanish in the local Lorentz frame at point $\mathscr{P}_0$—and required to vanish by the *physics*.) From $\boldsymbol{\nabla g} = 0$, one can derive both the abstract chain rule

$$\boldsymbol{\nabla}_{\boldsymbol{u}}(\boldsymbol{v}\cdot\boldsymbol{w}) = (\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v})\cdot\boldsymbol{w} + \boldsymbol{v}\cdot(\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}) \tag{13.20}$$

$\Gamma^{\alpha}{}_{\beta\gamma}$ expressed in terms of metric

(Exercise 13.4) and the following equations for the connection coefficients in *any* frame in terms of (1) the metric coefficients, $g_{\alpha\beta} = \boldsymbol{e}_\alpha\cdot\boldsymbol{e}_\beta$, and (2) the covariant commutation coefficients

$$c_{\alpha\beta\gamma} \equiv c_{\alpha\beta}{}^{\mu} g_{\mu\gamma} \equiv \langle \boldsymbol{\omega}^{\mu}, [\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta]\rangle g_{\mu\gamma} \tag{13.21}$$

of that frame:

$$\Gamma^{\alpha}{}_{\beta\gamma} = g^{\alpha\mu}\Gamma_{\mu\beta\gamma} \qquad (\text{definition of } \Gamma_{\mu\beta\gamma}), \tag{13.22}$$

$$\begin{aligned}\Gamma_{\mu\beta\gamma} &= \frac{1}{2}(g_{\mu\beta,\gamma} + c_{\mu\beta\gamma} + g_{\mu\gamma,\beta} + c_{\mu\gamma\beta} - g_{\beta\gamma,\mu} - c_{\beta\gamma\mu})\\ &= \frac{1}{2}(g_{\mu\beta,\gamma} + g_{\mu\gamma,\beta} - g_{\beta\gamma,\mu}) \text{ in any coordinate frame.}\end{aligned} \tag{13.23}$$

(See Exercise 13.4).

Equations (13.23) are the connection coefficients required to make the geodesics of curved spacetime coincide with the straight lines of the local Lorentz geometry. And they are fixed uniquely; no other choice of connection coefficients will do the job!

Summary: in curved spacetime with a local Lorentz metric, the following seemingly different statements are actually equivalent: (1) the geodesics of curved spacetime coincide with the straight lines of the local Lorentz geometry; (2) every local Lorentz frame [coordinates with $g_{\alpha\beta}(\mathscr{P}_0) = \eta_{\alpha\beta}$, $g_{\alpha\beta,\gamma}(\mathscr{P}_0) = 0$] is a local inertial frame [$\Gamma^{\alpha}{}_{\beta\gamma}(\mathscr{P}_0) = 0$]; (3) the metric and covariant derivative satisfy the compatibility condition $\boldsymbol{\nabla g} = 0$; (4) the covariant derivative obeys the chain rule (13.20); (5) the connection coefficients are determined by the metric in the manner of equations (13.23). A sixth equivalent statement, derived in the next section, says (6) the geodesics of curved spacetime coincide with world lines of extremal proper time.

EXERCISES

Exercise 13.3. MATHEMATICAL REPRESENTATION OF LOCAL LORENTZ FRAME

By definition, a local Lorentz frame at a given event $\mathscr{P}_0$ is the closest thing there to a global Lorentz frame. Thus, it should be a coordinate system with $g_{\mu\nu}(\mathscr{P}_0) = \eta_{\mu\nu}$, and with as many derivatives of $g_{\mu\nu}$ as possible vanishing at $\mathscr{P}_0$. Prove that there exist coordinates in which $g_{\mu\nu}(\mathscr{P}_0) = \eta_{\mu\nu}$ and $g_{\mu\nu,\rho}(\mathscr{P}_0) = 0$, but that $g_{\mu\nu,\rho\sigma}(\mathscr{P}_0)$ cannot vanish in general. Hence, such coordinates are the mathematical representation of a local Lorentz frame. [Hint: Let $\{x^{\alpha'}(\mathscr{P})\}$ be an arbitrary but specific coordinate system, and $\{x^{\mu}(\mathscr{P})\}$ be a local Lorentz frame, both with origins at $\mathscr{P}_0$. Expand the coordinate transformation between the two in powers of x^μ

$$x^{\alpha'} = M^\alpha{}_\mu x^\mu + \frac{1}{2} N^\alpha{}_{\mu\nu} x^\mu x^\nu + \frac{1}{6} P^\alpha{}_{\mu\nu\rho} x^\mu x^\nu x^\rho + \dots;$$

and use the transformation matrix $L^{\alpha'}{}_\mu \equiv \partial x^{\alpha'}/\partial x^\mu$ to get $g_{\mu\nu}(\mathscr{P}_0)$, $g_{\mu\nu,\rho}(\mathscr{P}_0)$, and $g_{\mu\nu,\rho\sigma}(\mathscr{P}_0)$ in terms of $g_{\alpha'\beta'}$ and its derivatives and the constants $M^\alpha{}_\mu$, $N^\alpha{}_{\mu\nu}$, $P^\alpha{}_{\mu\nu\rho}$. Show that whatever $g_{\alpha'\beta'}$ may be (so long as it is nonsingular, so $g^{\alpha'\beta'}$ exists!), one can choose the 16 constants $M^\alpha{}_\mu$ to make $g_{\mu\nu} = \eta_{\mu\nu}$ (ten conditions); one can choose the $4 \times 10 = 40$ constants $N^\gamma{}_{\mu\nu}$ to make the $10 \times 4 = 40$ $g_{\mu\nu,\rho}(\mathscr{P}_0)$ vanish; but one cannot in general choose the $4 \times 20 = 80$ $P^\alpha{}_{\mu\nu\rho}$ to make the $10 \times 10 = 100$ $g_{\mu\nu,\rho\sigma}$ vanish.]

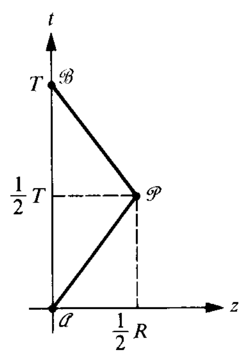
Exercise 13.4. CONSEQUENCES OF COMPATIBILITY BETWEEN g AND ∇

(a) From the condition of compatibility $\nabla g = 0$, derive the chain rule (13.20).

(b) From the condition of compatibility $\nabla g = 0$ and definitions (13.21) and (13.22), derive equation (13.23) for the connection coefficients. [Answer: See exercise 8.15, p. 216.]

---

§13.4. GEODESICS AS WORLD LINES OF EXTREMAL PROPER TIME



In a local Lorentz frame, it is easy to distinguish a world line that is straight from one that is not. Position the Lorentz frame and so orient it that the starting point of the world line, $\mathscr{A}$, lies at the origin and the end point, $\mathscr{B}$, lies at $x = 0$, $y = 0$, $z = 0$, $t = T$. As an example of a nonstraight world line, consider passage at uniform velocity from $\mathscr{A}$ to point $\mathscr{P}$ with coordinates $(\frac{1}{2}T; 0, 0, \frac{1}{2}R)$ and from there again with uniform velocity to point $\mathscr{B}$. The lapse of proper time from start to finish ("length of world line") is

$$\tau = (T^2 - R^2)^{1/2}.$$

Thus the lapse of proper time is diminished from its straight-line value, and diminished moreover for any choice of R whatsoever, except for the zero or straight-line value $R = 0$. As for this simple nonstraight curve, so also for any other nonstraight curve: the lapse of proper time between $\mathscr{A}$ and $\mathscr{B}$ is less than the straight-line lapse (Exercise 6.3). Thus, in flat spacetime, extremal length of world line is an indicator of straightness.

In flat spacetime, straight lines have extremal length

Any local region of the curved spacetime of the real, physical world is Lorentz in character. In this local Lorentz geometry, it is easy to set up Lorentz coordinates and carry out the extremal-length analysis just sketched to distinguish between a straight line and a nonstraight line:

Extremal length in curved spacetime

$$\tau = \int_{\mathscr{A}}^{\mathscr{B}} d\tau = \int_{\mathscr{A}}^{\mathscr{B}} (-\eta_{\mu\nu}\, dx^\mu\, dx^\nu)^{1/2}$$
$$= \begin{pmatrix} \text{a maximum for straight line} \\ \text{as compared to any variant of} \\ \text{the straight line} \end{pmatrix}. \qquad (13.24)$$

Such a test for straightness can be carried out separately in each local Lorentz region along the world line, or, with greater efficiency, it can be carried out over many local Lorentz regions simultaneously, i.e., over a region with endpoints $\mathcal{A}$ and $\mathcal{B}$ so widely separated that no single Lorentz frame can possibly contain them both. To carry out the analysis, one must abandon local Lorentz coordinates. Therefore introduce a general curvilinear coordinate system and find

$$\tau = \int_{\mathcal{A}}^{\mathcal{B}} d\tau = \int_{\mathcal{A}}^{\mathcal{B}} (-g_{\mu\nu}\, dx^\mu\, dx^\nu)^{1/2} = \begin{pmatrix} \text{an extremum for timelike world line that} \\ \text{is straight in each local Lorentz frame} \\ \text{along its path, as compared to any "nearby"} \\ \text{variant of this world line} \end{pmatrix}. \tag{13.25}$$

In the real world, the path of extremal τ, being straight in every local Lorentz frame, must be a geodesic of spacetime.

Notice that the word "maximum" in equation (13.24) has been replaced by "extremum" in the statement (13.25). When $\mathcal{A}$ and $\mathcal{B}$ are widely separated, they may be connected by several different geodesics with differing lapses of proper time (Figure 13.2). Each timelike geodesic extremizes τ with respect to nearby deformations of itself, but the extremum need not be a maximum. When several distinct geodesics connect two events, the typical one is not a local maximum ("mountain peak") but a saddle point ("mountain pass") in such a diagram as Figure 13.2 or 13.3.

Proof that curves of extremal length are geodesics

Concord between locally straight lines (lines of extremal τ) and geodesics of curved spacetime demands that timelike geodesics have extremal proper length. If so, then any curve $x^\mu(\lambda)$ between $\mathcal{A}$ (where $\lambda = 0$) and $\mathcal{B}$ (where $\lambda = 1$) that extremizes τ should satisfy the geodesic equation. To test for an extremal by comparing times, pick a curve suspected to be a geodesic, and deform it slightly but arbitrarily:

$$\begin{aligned} &\text{original curve, } x^\mu = a^\mu(\lambda); \\ &\text{deformed curve, } x^\mu = a^\mu(\lambda) + \delta a^\mu(\lambda). \end{aligned} \tag{13.26}$$

Along either curve the lapse of proper time is

$$\tau = \int_{\mathcal{A}}^{\mathcal{B}} d\tau = \int_0^1 \left(-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\right)^{1/2} d\lambda. \tag{13.27}$$

At fixed λ the metric coefficient $g_{\mu\nu}[x^\alpha(\lambda)]$ differs from one curve to the other by

$$\delta g_{\mu\nu} \equiv g_{\mu\nu}[a^\alpha(\lambda) + \delta a^\alpha(\lambda)] - g_{\mu\nu}[a^\alpha(\lambda)] = \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\,\delta a^\sigma(\lambda); \tag{13.28}$$

and the components $dx^\nu/d\lambda$ of the tangent vector differ by

$$\delta\left(\frac{dx^\nu}{d\lambda}\right) \equiv \frac{d(a^\nu + \delta a^\nu)}{d\lambda} - \frac{da^\nu}{d\lambda} = \frac{d}{d\lambda}(\delta a^\nu). \tag{13.29}$$

These changes in $g_{\mu\nu}$ and $dx^\nu/d\lambda$, at fixed λ, produce corresponding changes in the lapse of proper time in equation (13.27):

$$\delta\tau = \int_0^1 \left\{ \frac{-g_{\mu\nu}(da^\mu/d\lambda)d(\delta a^\nu)/d\lambda - \frac{1}{2}(g_{\mu\nu,\sigma}\delta a^\sigma)(da^\mu/d\lambda)(da^\nu/d\lambda)}{[-g_{\gamma\delta}(da^\gamma/d\lambda)(da^\delta/d\lambda)]^{1/2}} \right\} d\lambda.$$

Integrate the first term by parts. Strike out the end-point terms, because both paths must pass through $\mathscr{A}$ and $\mathscr{B}$ ($\delta a^\mu = 0$ at $\lambda = 0$ and $\lambda = 1$). Thus find

$$\delta\tau = \int_{\lambda=0}^{\lambda=1} f_\sigma(\lambda)\,\delta a^\sigma \left[-g_{\gamma\delta}\frac{da^\gamma}{d\lambda}\frac{da^\delta}{d\lambda} \right]^{1/2} d\lambda. \tag{13.30}$$

Here the f_σ ("force terms") in the integrand are abbreviations for the four expressions

$$f_\sigma(\lambda) = \frac{1}{\left[-g_{\gamma\delta}\frac{da^\gamma}{d\lambda}\frac{da^\delta}{d\lambda} \right]^{1/2}} \frac{d}{d\lambda} \frac{g_{\sigma\nu}\frac{da^\nu}{d\lambda}}{\left[-g_{\gamma\delta}\frac{da^\gamma}{d\lambda}\frac{da^\delta}{d\lambda} \right]^{1/2}} - \frac{\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\frac{da^\mu}{d\lambda}\frac{da^\nu}{d\lambda}}{\left[-g_{\gamma\delta}\frac{da^\gamma}{d\lambda}\frac{da^\delta}{d\lambda} \right]}. \tag{13.31}$$

An extremum is achieved, and the first-order change $\delta\tau$ vanishes for every first-order deformation $\delta a^\sigma(\lambda)$ from an optimal path $x^\sigma = a^\sigma(\lambda)$, when the four quantities f_σ that multiply the δa^σ all vanish. Thus one arrives at the four conditions

$$f_\sigma(\lambda) = 0 \tag{13.32}$$

for the determination of an extremal world line. (An alternative viewpoint on the extremization is spelled out in Figure 13.3.)

Sufficient these four equations are, but *independent* they are not, by reason of a "bead argument" (automatic vanishing of $\delta\tau$ for any set of changes that merely slide points, like beads, along an existing world line). The operation of mere "sliding of beads" implies the trivial change

$$\delta a^\sigma(\lambda) = h(\lambda)\frac{da^\sigma}{d\lambda}, \tag{13.33}$$

where $h(\lambda)$ is an arbitrary function of position along the world line ("more sliding here than there"). Already knowing that this operation cannot change τ, one is guaranteed that the integrand in (13.30) must vanish when one inserts (13.33) for δa^σ; and must vanish, moreover, whatever choice is made for the arbitrary "magnitude of slide" factor $h(\lambda)$. This requirement implies and demands that the scalar product $f_\sigma\, da^\sigma/d\lambda$ must automatically vanish; or, otherwise stated,

$$f_\sigma \frac{da^\sigma}{d\tau} = 0. \tag{13.34}$$

The argument applies, and this equation holds, whether one is or is not dealing with an optimal world line. An equation of this type, valid whether or not the world line is an allowable track for a free test particle (track of extremal lapse of proper time), is known as an *identity*. Equation (13.34), an important identity in the realm of spacetime geodesics, is an appropriate forerunner for the Bianchi identities of Chapter 15: the most important identities in the realm of spacetime curvature.

The freedom that exists to "slide λ-values along the world line" can be exploited to replace the arbitrary parameter λ by the physically more interesting parameter of proper time itself,

$$d\tau = \left[-g_{\gamma\delta} \frac{da^\gamma}{d\lambda} \frac{da^\delta}{d\lambda} \right]^{1/2} d\lambda. \tag{13.35}$$

Figure 13.2.
Star oscillating back and forth through the plane of a disc galaxy, as an example of a situation where two events $\mathcal{A}$ and $\mathcal{B}$ can be connected by more than one geodesic. Upper left: The galaxy seen edge-on, showing (dashed line) the path of the star in question, referred to a local frame partaking of and comoving with the general revolution of the nearby "disc stars." Upper right: The effective potential sensed by the star, according to Newtonian gravitation theory, is like that experienced by a ball which rolls down one inclined plane and up another ("free fall toward galactic plane" with acceleration $g = \frac{1}{2}$ in the units used here). The three central frames: Possible and impossible world lines for the star connecting two given events $\mathcal{A}$ (plane of galaxy at $t = 0$) and $\mathcal{B}$ (plane of galaxy at $t = 2$). Right: Throw star up from the galactic plane with enough velocity so that it just gets back to the plane at $t = 2$. Left: Throw it up with half the velocity and it will come back in half the time (very contrary to behavior of a simple harmonic oscillation, but in accord with galaxy's v-shaped potential!), thus being able to make two excursions in the allotted time between $\mathcal{A}$ and $\mathcal{B}$. Center: A conceivable world line (conceivable with rocket propulsion!) but not a geodesic. Bottom: Comparison of these and any other paths that allow themselves to be approximated in the form

$$z = a_1 \sin(\pi t/2) + a_2 \sin(2\pi t/2).$$

Here the two adjustable parameters, a_1 and a_2, provide the coordinates in a two-dimensional "function space" (approximation to the infinite-dimensional function space required to depict all conceivable world lines connecting $\mathcal{A}$ and $\mathcal{B}$; note comparison in right center frame between one-term Fourier approximation and exact, parabolic law of free fall; similarly in left center frame, where the two curves agree too closely to be shown separate on the diagram). Details: In the context of general relativity, take an arbitrary world line that connects $\mathcal{A}$ and $\mathcal{B}$, evaluate lapse of proper time, repeat for other world lines, and say that a given world line represents a possible motion ("geodesic") when for it the proper time is an extremum with respect to all nearby world lines. In the Newtonian approximation, the difference between the lapse of proper time and the lapse $(t_\mathcal{B} - t_\mathcal{A})$ of coordinate time is all that comes to attention, in the form of the "action integral" (on a "per-unit-mass basis")

$$I = \int_\mathcal{A}^\mathcal{B} \left[\begin{pmatrix}\text{kinetic}\\ \text{energy}\end{pmatrix} - \begin{pmatrix}\text{potential}\\ \text{energy}\end{pmatrix} \right] dt$$
$$= \int \left[\frac{1}{2}\left(\frac{dz}{dt}\right)^2 - |z| \right] dt$$

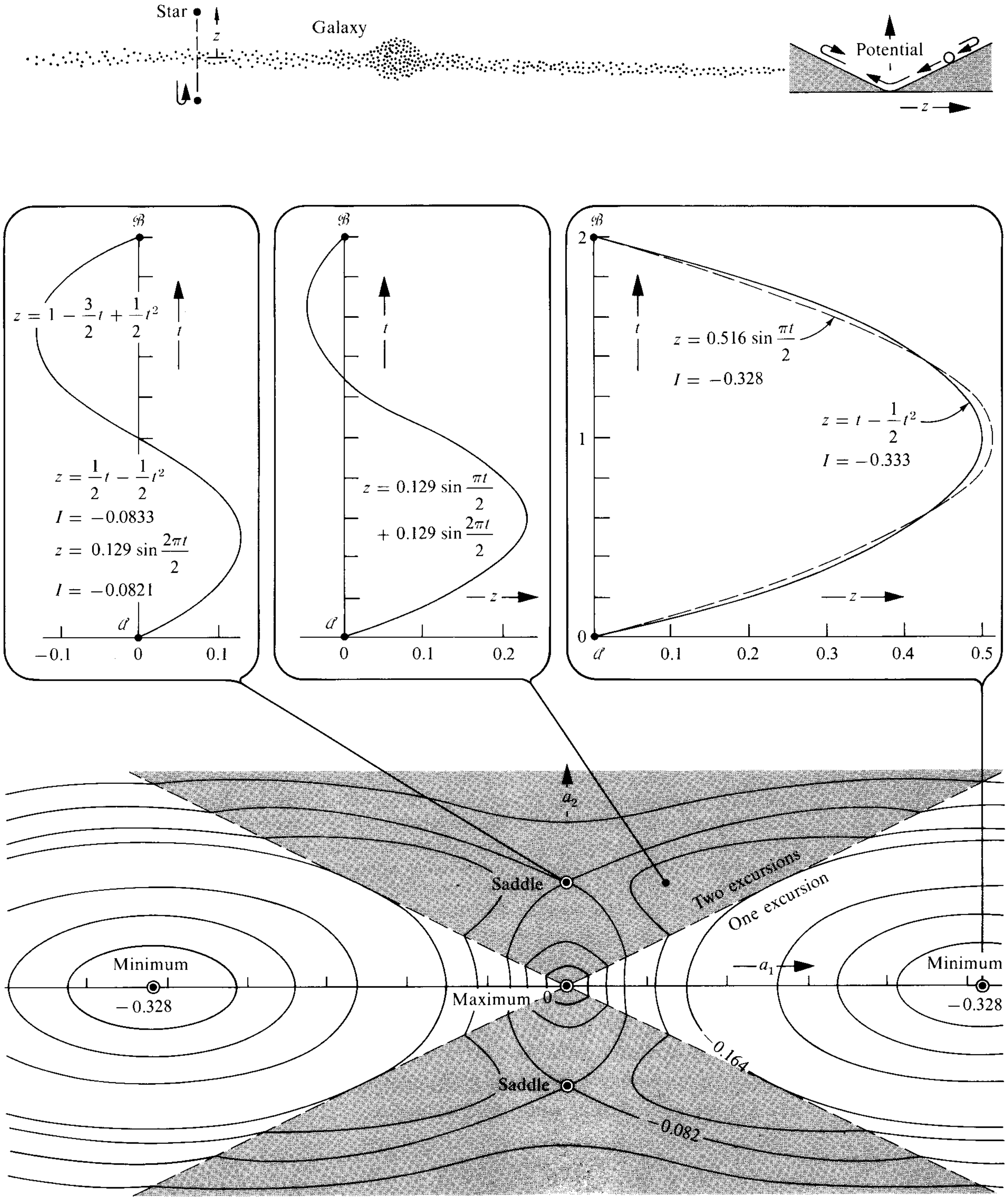
(maximum, or other extremum, in the proper time implies minimum, or corresponding other extremum, in the action I). The integration gives

$$I = (\pi^2 a_1^2/8) - (4|a_1|/\pi) + (\pi^2 a_2^2/2)$$

for $|a_2| < \frac{1}{2}|a_1|$ (one-excursion motions), and for $|a_2| > \frac{1}{2}|a_1|$ (two-excursion motions),

$$I = (\pi^2 a_1^2/8) + (\pi^2 a_2^2/2) - (4|a_2|/\pi) - (a_1^2/\pi|a_2|).$$

The one-excursion motion minimizes the action (maximizes the lapse of proper time). The two-excursion motion extremizes the action but does not minimize it ("saddle point"; "mountain pass" in the topography). Choquard (1955) gives other examples of problems of mechanics where there is more than one extremum. Morse (1934) and Morse and Cairns (1969) give a theorem connecting the number of saddles of various types with the numbers of maxima and minima ("critical-point theorem of the calculus of variations in the large").


Star
Galaxy
z
Potential
— z →
$\mathscr{B}$
$z = 1 - \frac{3}{2}t + \frac{1}{2}t^2$
t
$z = \frac{1}{2}t - \frac{1}{2}t^2$
$I = -0.0833$
$z = 0.129 \sin \frac{2\pi t}{2}$
$I = -0.0821$
$\mathscr{A}$
−0.1
0
0.1
$z = 0.129 \sin \frac{\pi t}{2}$
$+ 0.129 \sin \frac{2\pi t}{2}$
0.2
2
1
$z = 0.516 \sin \frac{\pi t}{2}$
$I = -0.328$
$z = t - \frac{1}{2}t^2$
$I = -0.333$
0.3
0.4
0.5
a_2
Saddle
Two excursions
One excursion
Minimum
− 0.328
Maximum 0
— a_1 →
−0.164
−0.082

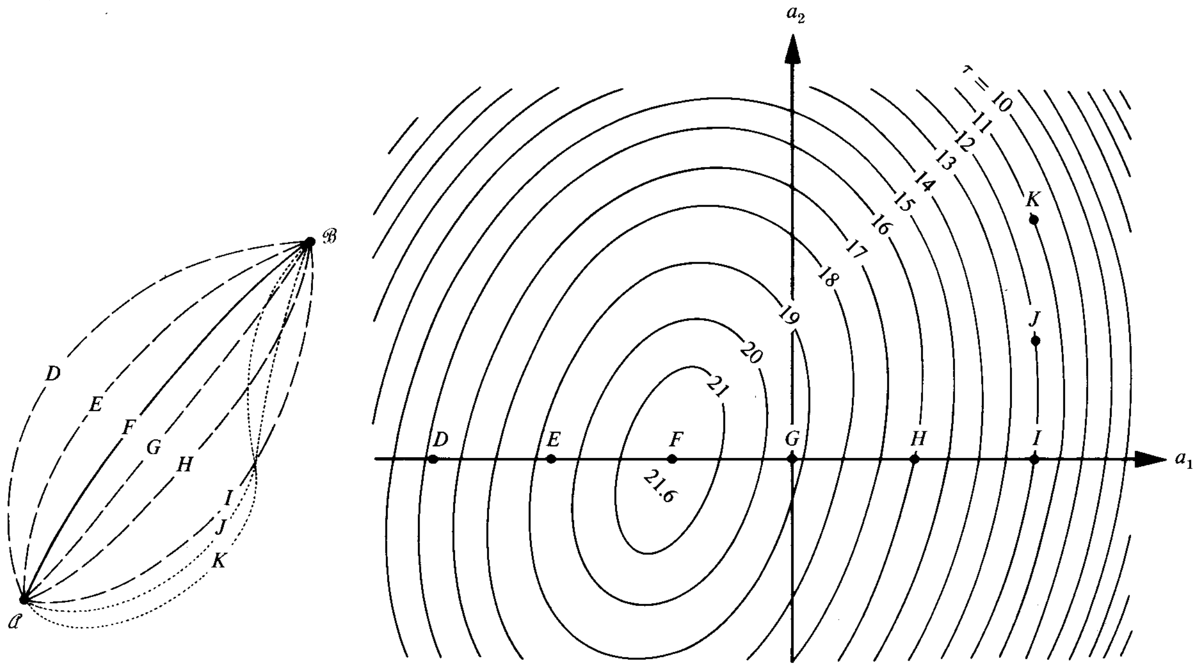


Figure 13.3.
Extremizing lapse of proper time by suitable choice of world line. Left: Spacetime; and world line F that extremizes the lapse of proper time τ from $\mathscr{A}$ to $\mathscr{B}$ compared to other world lines. The specific world lines depicted in the diagram happen to be distinguished from fiducial world line G by two "Fourier amplitudes" a_1 and a_2:

$$\delta a^\mu(\lambda) = a_1 \sin(\pi\lambda) + a_2 \sin(2\pi\lambda),$$

where the arbitrary scaling of λ, and its zero, are so adjusted that $\lambda(\mathscr{A}) = 0$, $\lambda(\mathscr{B}) = 1$.

Right: "Path space." The coordinates in this space are the Fourier amplitudes a_1 and a_2. Only these two amplitudes ("two dimensions") are shown out of what in principle are infinitely many amplitudes ("infinite-dimensional path space") required to represent the general timelike world line connecting $\mathscr{A}$ and $\mathscr{B}$. Any given contour curve runs through all those points (in path space) for which the corresponding world lines (in spacetime) rack up the indicated lapse of proper time τ. Foregoing description is classical; according to quantum mechanics, all the timelike world lines connecting $\mathscr{A}$ and $\mathscr{B}$ occur with the same probability amplitude ("principle of democracy of histories") with the only difference from one to another being the phase of this complex probability amplitude $\exp(-im\tau/\hbar)$ (m = mass of particle, $\hbar$ = quantum of angular momentum). In the sum over these probability amplitudes, however, destructive interference wipes out the contributions from all those histories which differ too much from the optimal or classical history ("Fresnel wave zone"; "Feynman's principle of sum over histories"; see Feynman and Hibbs, 1965). Capitalizing on this wave-mechanical background to show how the machinery of the physical world works, Box 25.3 spells out the Hamilton-Jacobi method ("short-wavelength limit of quantum mechanics") for determining geodesics, a method considerably more convenient for most applications than the usual "second-order differential equations for geodesics" (equation 10.27).

Focus on a specific world line, $x^\mu = a^\mu(\lambda)$, with all deformations of it gone from view; one may replace $a^\mu(\lambda)$ by $x^\mu(\lambda)$ everywhere. Then the differential equations (13.32) for an extremal world line reduce to

$$g_{\sigma\nu}\frac{d^2x^\nu}{d\tau^2} + \frac{1}{2}\left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\sigma\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right)\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0. \qquad (13.36)$$

As an aside, note that the identity (13.34) now follows by one differentiation (with respect to τ) of the equation

$$g_{\sigma\nu}\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau}+1=0. \tag{13.37}$$

Thus the identity is to be interpreted as saying that 4-velocity and 4-acceleration are orthogonal for any world line, extremal or not. Now return to (13.36), raise an index with $g^{\beta\sigma}$, and thereby bring the equation for a straight line of local Lorentz geometry into the form

$$\frac{d^2x^\beta}{d\tau^2}+g^{\beta\sigma}\frac{1}{2}\left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu}+\frac{\partial g_{\sigma\mu}}{\partial x^\nu}-\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right)\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}=0. \tag{13.38}$$

Compare with the standard form of the equation for a geodesic in "premetric geometry,"

$$\frac{d^2x^\beta}{d\lambda^2}+\Gamma^\beta{}_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}=0. \tag{13.39}$$

Conclude that the geodesics of the premetric geometry will agree with the straight lines of the local Lorentz geometry if and only if two conditions are satisfied: (1) the 40 connection coefficients $\Gamma^\beta{}_{\mu\nu}$ that define geodesics, covariant derivatives, and parallel transport must be given in terms of the 10 metric coefficients $g_{\mu\nu}$ ("Einstein gravitation potentials") by the equations (13.22) and (13.23) previously derived; and (2) the geodesic parameter λ must agree with the proper time τ up to an arbitrary normalization of zero point and an arbitrary but constant scale factor; thus

$$\lambda = a\tau + b.$$

(Nothing in the formalism has any resemblance whatsoever to the universal time t of Newton "flowing everywhere uniformly"; rather, there is a separate proper time τ for each geodesic). See Box 13.3 for another variational principle, which gives in one step both the extremal world line and the right parametrization on that line.

With this step, one has completed the transfer of the ideas of curved-space geometry from a foundation based on geodesics to a foundation based on metric. The resulting geometry always and everywhere anchors itself to the principle of "local Lorentz character," as the geometry of Newton-Cartan never did and never could.

EXERCISES

Exercise 13.5. ONCE TIMELIKE, ALWAYS TIMELIKE

Show that a geodesic of spacetime which is timelike at one event is everywhere timelike. Similarly, show that a geodesic initially spacelike is everywhere spacelike, and a geodesic initially null is everywhere null. [*Hint:* This is the easiest exercise in the book!]

(*continued on page 324*)

Box 13.2 "GEODESIC" VERSUS "EXTREMAL WORLD LINE"

Once the connection coefficients $\Gamma^{\alpha}{}_{\mu\nu}$ have been expressed in terms of Einstein's gravitational potentials $g_{\mu\nu}$ by the equations (13.22) and (13.23), as they are now and hereafter will be in this book ("Riemannian or metric geometry"), it is permissible and appropriate to subsume under the one word "geodesic" two previously distinct ideas: (1) a parametrized world line that satisfies the geodesic equation

$$\frac{d^2x^{\alpha}}{d\lambda^2} + \Gamma^{\alpha}{}_{\mu\nu}\frac{dx^{\mu}}{d\lambda}\frac{dx^{\nu}}{d\lambda} = 0;$$

and (2) a world line that extremizes the proper time (or, if spacelike, a curve that extremizes the proper distance) between two events $\mathcal{A}$ and $\mathcal{B}$. The one possible source of confusion is this, that (1) presupposes a properly parametrized curve (as was essential, for example, in the Schild's ladder construction employed for parallel transport in Chapter 10), whereas (2) cares only about the course of the world line through spacetime, being indifferent to what parametrization is used or whether any parametrization at all is introduced. This is not to deny the possibility of "marking in afterward" along the extremal curve the most natural and easily evaluated of all parameters, the proper time itself, whereupon the extremal curve of (2) satisfies the geodesic equation of (1). Ambiguity is avoided by insisting on proper parametrization: henceforth the word "curve" means a parametrized curve, the word "geodesic" means a properly parametrized geodesic.

Box 13.3 "DYNAMIC" VARIATIONAL PRINCIPLE FOR GEODESICS

If the principle of extremal length

$$\tau = \int_{\mathcal{A}}^{\mathcal{B}} \left[-g_{\mu\nu}\frac{dx^{\mu}}{d\lambda}\frac{dx^{\nu}}{d\lambda}\right]^{1/2} d\lambda = \text{extremum} \qquad (1)$$

is indifferent to choice of parametrization ["$d\lambda$" canceling out in (1)] and if the geodesic equation finds the proper parametrization a matter of concern, it is appropriate to search for another extremal principle that yields in one package both the right curve and the right parameter. By analogy with elementary mechanics, one expects that an equation of motion [the geodesic equation

$$d^2x^{\mu}/d\lambda^2 + \Gamma^{\mu}{}_{\alpha\beta}\,(dx^{\alpha}/d\lambda)(dx^{\beta}/d\lambda) = 0]$$

whose leading term has the form "$\ddot{x}$" can be derived from a Lagrangian with leading term "$\frac{1}{2}\dot{x}^2$" ("kinetic energy"; "dynamic" term). The simplest coordinate invariant generalization of $\frac{1}{2}\dot{x}^2$ is

$$\frac{1}{2}\, g_{\mu\nu}\,(dx^{\mu}/d\lambda)(dx^{\nu}/d\lambda).$$

Thus one is led to try, in place of the "geometric" principle of extremal length, a new "dynamic" extremal principle:

$$\begin{aligned} I &= \frac{1}{2}\int_{\mathcal{A}}^{\mathcal{B}} g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\,d\lambda \\ &= \int_{\mathcal{A}}^{\mathcal{B}} L\left(x^\sigma, \frac{dx^\sigma}{d\lambda}\right)d\lambda = \text{extremum} \end{aligned} \tag{2}$$

(replacement of square root in previous variational principle by first power). The condition for an extremum, here as before [equations (13.30) to (13.32)] is annulment of the so-called Euler-Lagrange "functional derivative"

$$\begin{aligned} 0 = \frac{\delta I}{\delta x^\sigma} &\equiv \begin{pmatrix}\text{coefficient of } \delta x^\sigma \text{ in} \\ \text{the integrand of } \delta I\end{pmatrix} \\ &= \frac{\partial L}{\partial x^\sigma} - \frac{d}{d\lambda}\frac{\partial L}{\partial\left(\dfrac{dx^\sigma}{d\lambda}\right)}; \end{aligned} \tag{3}$$

or, written out in full detail,

$$g_{\sigma\nu}\frac{d^2x^\nu}{d\lambda^2} + \frac{1}{2}\left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\sigma\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right)\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0; \tag{4}$$

or, after multiplication by the reciprocal metric,

$$\frac{d^2x^\alpha}{d\lambda^2} + g^{\alpha\sigma}\frac{1}{2}\left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\sigma\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right)\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0; \tag{5}$$

which translates into the geodesic equation

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha{}_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0. \tag{6}$$

Thus, the new "dynamic" expression (2) is indeed extremal for geodesic curves—and, by contrast with proper length, (1), it is extremal when and only when the geodesic is affinely parametrized. [Its "Euler-Lagrange equations" (6) remain satisfied only under parameter changes $\lambda_{\text{new}} = a\lambda_{\text{old}} + b$, which keep the parameter affine; by contrast, the Euler-Lagrange equations (13.31) and (13.32) for the "principle of extremal length" (1) remain satisfied for any change of parameter whatsoever.]

Exercise 13.6. SPACELIKE GEODESICS HAVE EXTREMAL LENGTH

Show that any spacelike curve linking two events $\mathscr{A}$ and $\mathscr{B}$ is a geodesic if and only if it extremizes the proper length

$$s = \int_{\mathscr{A}}^{\mathscr{B}} (g_{\mu\nu}\, dx^\mu\, dx^\nu)^{1/2}.$$

[*Hint:* This is almost as easy as exercise 13.5 if one has already proved the analogous theorem for timelike geodesics.]

Exercise 13.7. METRIC TENSOR MEASURED BY LIGHT SIGNALS AND FREE PARTICLES [Kuchař]

(a) Instead of parametrizing a timelike geodesic by the proper time τ, parametrize it by an arbitrary parameter μ,

$$\tau = F(\mu).$$

Write the geodesic equation in the μ-parametrization.

(b) Use now the coordinate time t as a parameter. Throw out a cloud of free particles with different "velocities" $v^i = dx^i/dt$ and observe their "accelerations" $a^i = d^2x^i/dt^2$. Discuss what combinations of the components of the affine connection $\Gamma^\iota{}_{\kappa\lambda}$ one can measure in this way. (Assume that no standard clocks measuring τ are available!)

(c) Show that one can measure the conformal metric $\bar{g}_{\iota\kappa}$, i.e., the ratios of the components of the metric tensor $g_{\iota\kappa}$ to a given component (say, g_{00})

$$\bar{g}_{\iota\kappa} = \Lambda g_{\iota\kappa}, \qquad \Lambda \equiv (-g_{00})^{-1},$$

using only the light signals moving along the null geodesics $g_{\iota\kappa}\, dx^\iota\, dx^\kappa = 0$.

(d) Combine now the results of (b) and (c). Assume that $\Gamma^\iota{}_{\kappa\lambda}$ is generated by the metric tensor by (13.22), (13.23), in the coordinate frame x^ι. Show that one can determine Λ everywhere, if one prescribes it at one event (equivalent to fixing the unit of time).

---

§13.5. METRIC-INDUCED PROPERTIES OF *RIEMANN*

Symmetries of ***Riemann*** in absence of metric

In Newtonian spacetime, in the real, physical spacetime of Einstein—indeed, in any manifold with covariant derivative—the Riemann curvature tensor has these symmetries (exercise 11.6):

$$R^\alpha{}_{\beta\gamma\delta} \equiv R^\alpha{}_{\beta[\gamma\delta]} \qquad \text{(antisymmetry on last two indices)} \qquad (13.40)$$

$$R^\alpha{}_{[\beta\gamma\delta]} \equiv 0 \qquad \text{(vanishing of completely antisymmetric part).} \qquad (13.41)$$

In addition, it satisfies a differential identity (exercise 11.10):

$$R^\alpha{}_{\beta[\gamma\delta;\epsilon]} \equiv 0 \qquad \text{("Bianchi identity")} \qquad (13.42)$$

(see Chapter 15 for geometric significance).

New symmetries imposed by metric

When metric is brought onto the scene, whether in Einstein spacetime or elsewhere, it impresses on ***Riemann*** the additional symmetry (exercise 13.8)

$$R_{\alpha\beta\gamma\delta} \equiv R_{[\alpha\beta]\gamma\delta} \text{ (antisymmetry on first two indices).} \tag{13.43}$$

This, together with (13.40) and (13.41), forms a complete set of symmetries for ***Riemann***; other symmetries that follow from these (exercise 13.10) are

$$R_{\alpha\beta\gamma\delta} = R_{\gamma\delta\alpha\beta} \qquad \text{(symmetry under pair exchange),} \tag{13.44}$$

and

$$R_{[\alpha\beta\gamma\delta]} = 0 \qquad \text{(vanishing of completely antisymmetric part).} \tag{13.45}$$

These symmetries reduce the number of independent components of ***Riemann*** from $4 \times 4 \times 4 \times 4 = 256$ to 20 (exercise 13.9).

With metric present, one can construct a variety of new curvature tensors from ***Riemann***. Some that will play important roles later are as follows.

The curvature tensor $\boldsymbol{G}$

(1) The *double dual of* ***Riemann***, $\boldsymbol{G} \equiv$ ****Riemann**** (analog of ***Maxwell*** $\equiv$ ****Faraday***), which has components

$$G^{\alpha\beta}{}_{\gamma\delta} \equiv \frac{1}{2}\epsilon^{\alpha\beta\mu\nu} R_{\mu\nu}{}^{\rho\sigma} \frac{1}{2}\epsilon_{\rho\sigma\gamma\delta} = -\frac{1}{4}\delta^{\alpha\beta\mu\nu}_{\rho\sigma\gamma\delta} R_{\mu\nu}{}^{\rho\sigma} \tag{13.46}$$

(exercise 13.11).

Einstein tensor

(2) The *Einstein curvature tensor,* which is symmetric (exercise 13.11)

$$G^{\beta}{}_{\delta} \equiv G^{\mu\beta}{}_{\mu\delta}; \qquad G_{\beta\delta} = G_{\delta\beta}. \tag{13.47}$$

Ricci tensor

Curvature scalar

(3) The *Ricci curvature tensor,* which is symmetric, and the *curvature scalar*

$$R^{\beta}{}_{\delta} \equiv R^{\mu\beta}{}_{\mu\delta}, \qquad R_{\beta\delta} = R_{\delta\beta}; \qquad R \equiv R^{\beta}{}_{\beta}; \tag{13.48}$$

which are related to the Einstein tensor by (exercise 13.12)

$$R^{\beta}{}_{\delta} = G^{\beta}{}_{\delta} + \frac{1}{2} R \delta^{\beta}{}_{\delta}. \tag{13.49}$$

Weyl conformal tensor

(4) *The Weyl conformal tensor* (exercise 13.13)

$$C^{\alpha\beta}{}_{\gamma\delta} = R^{\alpha\beta}{}_{\gamma\delta} - 2\delta^{[\alpha}{}_{[\gamma} R^{\beta]}{}_{\delta]} + \frac{1}{3}\delta^{[\alpha}{}_{[\gamma}\delta^{\beta]}{}_{\delta]} R. \tag{13.50}$$

Bianchi identities

The Bianchi identity (13.42) takes a particularly simple form when rewritten in terms of the double dual $\boldsymbol{G}$:

$$G_{\alpha\beta\gamma}{}^{\delta}{}_{;\delta} \equiv 0 \qquad \text{("Bianchi identity")} \tag{13.51}$$

(exercise 13.11); and it has the obvious consequence

$$G_{\alpha}{}^{\beta}{}_{;\beta} \equiv 0 \qquad \text{("contracted Bianchi identity").} \tag{13.52}$$

Chapter 15 will be devoted to the deep geometric significance of these Bianchi identities.

EXERCISES

Exercise 13.8. *RIEMANN* ANTISYMMETRIC IN FIRST TWO INDICES

(a) Derive the antisymmetry condition (13.43). [Hint: Prove by abstract calculations that any vector fields $\boldsymbol{s}$, $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ satisfy $0 = \mathcal{R}(\boldsymbol{u}, \boldsymbol{v})(\boldsymbol{s} \cdot \boldsymbol{w}) = \boldsymbol{s} \cdot [\mathcal{R}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{w}] + \boldsymbol{w} \cdot [\mathcal{R}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{s}]$. Then from this infer (13.43).]

(b) Explain in geometric terms the meaning of this antisymmetry.

Exercise 13.9. NUMBER OF INDEPENDENT COMPONENTS OF *RIEMANN*

(a) In the absence of metric, a complete set of symmetry conditions for ***Riemann*** is $R^\alpha{}_{\beta\gamma\delta} = R^\alpha{}_{\beta[\gamma\delta]}$ and $R^\alpha{}_{[\beta\gamma\delta]} = 0$. Show that in four-dimensional spacetime these reduce the number of independent components from $4 \times 4 \times 4 \times 4 = 256$ to $4 \times 4 \times 6 - 4 \times 4 = 96 - 16 = 80$.

(b) Show that in a manifold of n dimensions without metric, the number of independent components is

$$\frac{n^3(n-1)}{2} - \frac{n^2(n-1)(n-2)}{6} = \frac{n^2(n^2-1)}{3}. \tag{13.53}$$

(c) In the presence of metric, a complete set of symmetries is $R_{\alpha\beta\gamma\delta} = R_{[\alpha\beta][\gamma\delta]}$, and $R_{\alpha[\beta\gamma\delta]} = 0$. Show that in four-dimensional spacetime, these reduce the number of independent components to $6 \times 6 - 4 \times 4 = 36 - 16 = 20$.

(d) Show that in a manifold of n dimensions with metric, the number of independent components is

$$\left[\frac{n(n-1)}{2}\right]^2 - \frac{n^2(n-1)(n-2)}{6} = \frac{n^2(n^2-1)}{12}. \tag{13.54}$$

Exercise 13.10. *RIEMANN* SYMMETRIC IN EXCHANGE OF PAIRS; COMPLETELY ANTISYMMETRIC PART VANISHES

From the complete set of symmetries in the presence of a metric, $R_{\alpha\beta\gamma\delta} = R_{[\alpha\beta][\gamma\delta]}$ and $R_{\alpha[\beta\gamma\delta]} = 0$, derive: (a) symmetry under pair exchange, $R_{\alpha\beta\gamma\delta} = R_{\gamma\delta\alpha\beta}$, and (b) vanishing of completely antisymmetric part, $R_{[\alpha\beta\gamma\delta]} = 0$. Then (c) show that the following form a complete set of symmetries:

$$R_{\alpha\beta\gamma\delta} = R_{[\alpha\beta][\gamma\delta]} = R_{\gamma\delta\alpha\beta}, \qquad R_{[\alpha\beta\gamma\delta]} = 0. \tag{13.55}$$

Exercise 13.11. DOUBLE DUAL OF *RIEMANN*; *EINSTEIN*

(a) Show that $\boldsymbol{G} \equiv$ ****Riemann**** contains precisely the same amount of information as ***Riemann***, and satisfies precisely the same set of symmetries [(13.40), (13.41), (13.43) to (13.45)].

(b) From the symmetries of $\boldsymbol{G}$, show that ***Einstein*** [defined in (13.47)] is symmetric ($G_{[\beta\delta]} = 0$).

(c) Show that the Bianchi identities (13.42), when written in terms of $\boldsymbol{G}$, take the form (13.51) ("vanishing divergence," $\boldsymbol{\nabla} \cdot \boldsymbol{G} = 0$).

(d) By contracting the Bianchi identities $\boldsymbol{\nabla} \cdot \boldsymbol{G} = 0$, show that $\boldsymbol{G} \equiv$ ***Einstein*** has vanishing divergence [equation (13.52)].

Exercise 13.12. *RICCI* AND *EINSTEIN* RELATED

(a) From the symmetries of ***Riemann***, show that ***Ricci*** is symmetric ($R_{[\beta\delta]} = 0$).

(b) Show that ***Ricci*** is related to ***Einstein*** by equation (13.49).

Exercise 13.13. THE WEYL CONFORMAL TENSOR

(a) Show that the Weyl conformal tensor (13.50) possesses the same symmetries [(13.40), (13.41), (13.43) to (13.45)] as the Riemann tensor.

(b) Show that the Weyl tensor is completely "trace-free"; i.e., that

$$\text{contraction of } C_{\alpha\beta\gamma\delta} \text{ on any pair of slots vanishes.} \tag{13.56}$$

Thus, $C_{\alpha\beta\gamma\delta}$ can be regarded as the trace-free part of ***Riemann***, and $R_{\alpha\beta}$ can be regarded as the trace of ***Riemann***. ***Riemann*** is determined entirely by its trace-free part $C_{\alpha\beta\gamma\delta}$ and its trace $R_{\alpha\beta}$ [see equation (13.50), and recall $R = R^{\alpha}{}_{\alpha}$].

(c) Show that in spacetime the Weyl tensor has 10 independent components.

(d) Show that in an n-dimensional manifold the number of independent components of ***Weyl*** [defined by a modification of (13.50) that maintains (13.56)] is

$$\frac{n^2(n^2-1)}{12} - \frac{n(n+1)}{2} \text{ for } n \geq 3, \qquad 0 \text{ for } n \leq 3. \tag{13.57}$$

Thus, in manifolds of 1, 2, or 3 dimensions, the Weyl tensor is identically zero, and the Ricci tensor completely determines the Riemann tensor.

§13.6. THE PROPER REFERENCE FRAME OF AN ACCELERATED OBSERVER

A physicist performing an experiment in a jet airplane (e.g., an infrared astronomy experiment) may use several different coordinate systems at once. But a coordinate system of special utility is one at rest relative to all the apparatus bolted into the floor and walls of the airplane cabin. This "proper reference frame" has a rectangular "$\hat{x}$, $\hat{y}$, $\hat{z}$" grid attached to the walls of the cabin, and one or more clocks at rest in the grid. That this proper reference frame is accelerated relative to the local Lorentz frames, the physicist knows from his own failure to float freely in the cabin, or, with greater precision, from accelerometer measurements. That his proper reference frame is rotating relative to local Lorentz frames he knows from the Coriolis forces he feels, or, with greater precision, from the rotation of inertial-guidance gyroscopes relative to the cabin walls.

Proper reference frame described physically

Exercise 6.8 gave a mathematical treatment of such an accelerated, rotating, but locally orthonormal reference frame in flat spacetime. This section does the same in curved spacetime. In the immediate vicinity of the spatial grid's origin $x^{\hat{j}} = 0$ (region of spatial extent so small that curvature effects are negligible), no aspect of the coordinate system can possibly reveal whether spacetime is curved or flat. Hence, all the details of exercise 6.8 must remain valid in curved spacetime. Nevertheless, it is instructive to rediscuss those details, and some new ones, using the powerful mathematics of the last few chapters.

Begin by making more precise the coordinate grid to be used. The following is perhaps the most natural way to set up the grid.

Six-step construction of coordinate grid for proper frame

(1) Let τ be proper time as measured by the accelerated observer's clock (clock at center of airplane cabin in above example). Let $\mathcal{P} = \mathcal{P}_0(\tau)$ be the observer's world line, as shown in Figure 13.4,a.

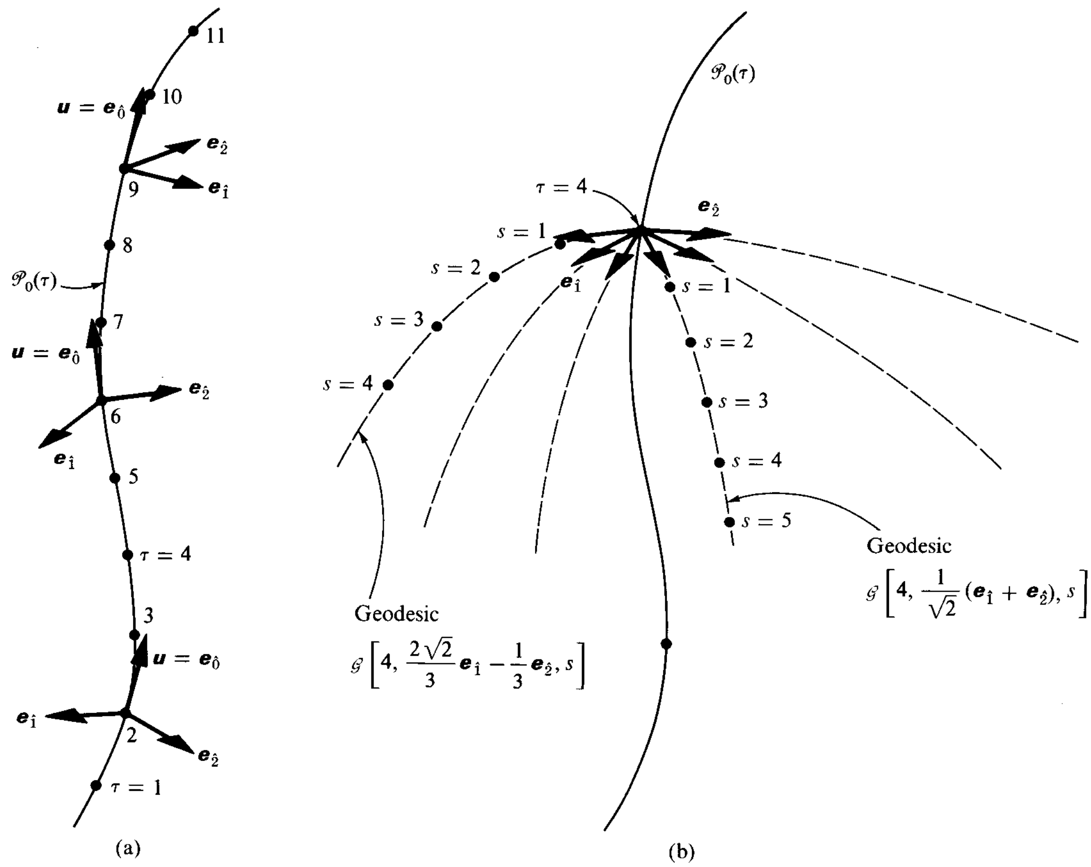


Figure 13.4.
The proper reference frame of an accelerated observer. Diagram (a) shows the observer's orthonormal tetrad $\{e_{\hat\alpha}\}$ being transported along his world line $\mathscr{P}_0(\tau)$ [transport law (13.60)]. Diagram (b) shows geodesics bristling out perpendicularly from an arbitrary event $\mathscr{P}_0(4)$ on the observer's world line. Each geodesic is specified uniquely by (1) the proper time τ at which it originates, and (2) the direction (unit tangent vector $\boldsymbol{n} = d/ds = n^{\hat j}e_{\hat j}$ along which it emanates). A given event on the geodesic is specified by τ, $\boldsymbol{n}$, and proper distance s from the geodesic's emanation point; hence the notation

$$\mathscr{P} = \mathscr{G}[\tau, \boldsymbol{n}, s]$$

for the given event. The observer's proper reference frame attributes to this given event the coordinates

$$x^{\hat 0}(\mathscr{G}[\tau, \boldsymbol{n}, s]) = \tau,$$
$$x^{\hat j}(\mathscr{G}[\tau, \boldsymbol{n}, s]) = sn^{\hat j}.$$

(2) The observer carries with himself an orthonormal tetrad $\{e_{\hat\alpha}\}$ (Figure 13.4,a), with

$$e_{\hat 0} = \boldsymbol{u} = d\mathscr{P}_0/d\tau = \text{(4-velocity of observer)} \tag{13.58}$$

($e_{\hat 0}$ points along observer's "time direction"), and with

$$e_{\hat\alpha} \cdot e_{\hat\beta} = \eta_{\alpha\beta} \tag{13.59}$$

(orthonormality).

(3) The tetrad changes from point to point along the observer's world line, relative to parallel transport:

Transport law for observer's tetrad

$$\nabla_{\boldsymbol{u}} \boldsymbol{e}_{\hat{\alpha}} = -\boldsymbol{\Omega} \cdot \boldsymbol{e}_{\hat{\alpha}}, \tag{13.60}$$

$$\begin{aligned} \Omega^{\mu\nu} &= a^\mu u^\nu - u^\mu a^\nu + u_\alpha \omega_\beta \epsilon^{\alpha\beta\mu\nu} \\ &= \text{"generator of infinitesimal Lorentz transformation."} \end{aligned} \tag{13.61}$$

This transport law has the same form in curved spacetime as in flat (§6.5 and exercise 6.8) because curvature can only be felt over finite distances, not over the infinitesimal distance involved in the "first time-rate of change of a vector" (equivalence principle). As in exercise 6.8,

$$\begin{aligned} \boldsymbol{a} &= \nabla_{\boldsymbol{u}} \boldsymbol{u} = (\text{4-acceleration of observer}), \\ \boldsymbol{\omega} &= \begin{pmatrix} \text{angular velocity of rotation of spatial} \\ \text{basis vectors } \boldsymbol{e}_{\hat{j}} \text{ relative to Fermi-} \\ \text{Walker-transported vectors, i.e.,} \\ \text{relative to inertial-guidance gyroscopes} \end{pmatrix}, \\ \boldsymbol{u} \cdot \boldsymbol{a} &= \boldsymbol{u} \cdot \boldsymbol{\omega} = 0. \end{aligned} \tag{13.62}$$

If $\boldsymbol{\omega}$ were zero, the observer would be Fermi-Walker-transporting his tetrad (gyroscope-type transport). If both $\boldsymbol{a}$ and $\boldsymbol{\omega}$ were zero, he would be freely falling (geodesic motion) and would be parallel-transporting his tetrad, $\nabla_{\boldsymbol{u}} \boldsymbol{e}_{\hat{\alpha}} = 0$.

(4) The observer constructs his proper reference frame (local coordinate system) in a manner analogous to the Riemann-normal construction of §11.6. From each event $\mathcal{P}_0(\tau)$ on his world line, he sends out purely spatial geodesics (geodesics orthogonal to $\boldsymbol{u} = d\mathcal{P}_0/d\tau$), with affine parameter equal to proper length,

$$\mathcal{P} = \mathcal{G}[\tau, \boldsymbol{n}, s]. \tag{13.63}$$

[proper time; tells "starting point" of geodesic]

[tangent vector to geodesic at starting point; tells "which" geodesic]

[proper length along geodesic from starting point; tells "where" on geodesic]

(See Figure 13.4,b.) The tangent vector has unit length, because the chosen affine parameter is proper length:

$$\begin{aligned} \boldsymbol{n} &= (\partial\mathcal{G}/\partial s)_{s=0}; \qquad n^\mu = (dx^\mu/ds) \text{ along geodesic,} \\ \boldsymbol{n} \cdot \boldsymbol{n} &= g_{\mu\nu}\left(\frac{dx^\mu}{ds}\right)\left(\frac{dx^\nu}{ds}\right) = \frac{ds^2}{ds^2} = 1. \end{aligned} \tag{13.64}$$

(5) Each event near the observer's world line is intersected by precisely one of the geodesics $\mathcal{G}[\tau, \boldsymbol{n}, s]$. [Far away, this is not true; the geodesics may cross, either because of the observer's acceleration, as in Figure 6.3, or because of the curvature of spacetime ("geodesic deviation").]

2

(6) Pick an event $\mathscr{P}$ near the observer's world line. The geodesic through it originated on the observer's world line at a specific time τ, had original direction $\boldsymbol{n} = n^{\hat{j}}\boldsymbol{e}_{\hat{j}}$, and needed to extend a distance s before reaching $\mathscr{P}$. Hence, the four numbers

$$(x^{\hat{0}}, x^{\hat{1}}, x^{\hat{2}}, x^{\hat{3}}) \equiv (\tau, sn^{\hat{1}}, sn^{\hat{2}}, sn^{\hat{3}}) \tag{13.65}$$

are a natural way of identifying the event $\mathscr{P}$. These are the coordinates of $\mathscr{P}$ in the observer's proper reference frame.

(7) Restated more abstractly,

$$\begin{aligned} x^{\hat{0}}(\mathscr{G}[\tau, \boldsymbol{n}, s]) &= \tau, \\ x^{\hat{j}}(\mathscr{G}[\tau, \boldsymbol{n}, s]) &= sn^{\hat{j}} = sn_{\hat{j}} = s\boldsymbol{n}\cdot\boldsymbol{e}_{\hat{j}}. \end{aligned} \tag{13.65$'$}$$

In flat spacetime this construction process and the resulting coordinates $x^{\hat{\alpha}}(\mathscr{P})$ are identical to the process and resulting coordinates $\xi^{\alpha'}(\mathscr{P})$ of exercise 6.8.

For use in calculations one wants not only the coordinate system, but also its metric coefficients and connection coefficients. Fortunately, $g_{\hat{\alpha}\hat{\beta}}$ and $\Gamma^{\hat{\alpha}}{}_{\hat{\beta}\hat{\gamma}}$ are needed only along the observer's world line, where they are especially simple. Only a foolish observer would try to use his own proper reference frame far from his world line, where its grid ceases to be orthonormal and its geodesic grid lines may even cross! (See §6.3.)

All along the observer's world line $\mathscr{P}_0(\tau)$, the basis vectors of his coordinate grid are identical (by construction) to his orthonormal tetrad

$$\partial/\partial x^{\hat{\alpha}} = \boldsymbol{e}_{\hat{\alpha}}, \tag{13.66}$$

and therefore its metric coefficients are

$$g_{\hat{\alpha}\hat{\beta}} = \boldsymbol{e}_{\hat{\alpha}}\cdot\boldsymbol{e}_{\hat{\beta}} = \eta_{\alpha\beta} \text{ all along } \mathscr{P}_0(\tau). \tag{13.67}$$

Connection coefficients along observer's world line

Some of the connection coefficients are determined by the transport law (13.60) for the observer's orthonormal tetrad:

$$\begin{aligned} \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{e}_{\hat{\alpha}} = \boldsymbol{\nabla}_{\hat{0}}\boldsymbol{e}_{\hat{\alpha}} &= \boldsymbol{e}_{\hat{\beta}}\Gamma^{\hat{\beta}}{}_{\hat{\alpha}\hat{0}} \\ &= -\boldsymbol{\Omega}\cdot\boldsymbol{e}_{\hat{\alpha}} = -\boldsymbol{e}_{\hat{\beta}}\Omega^{\hat{\beta}}{}_{\hat{\alpha}}. \end{aligned}$$

Thus

$$\Gamma^{\hat{\beta}}{}_{\hat{\alpha}\hat{0}} = -\Omega^{\hat{\beta}}{}_{\hat{\alpha}} \text{ all along } \mathscr{P}_0(\tau). \tag{13.68}$$

Since $\boldsymbol{\Omega}$ has the form (13.61) and the observer's 4-velocity and 4-acceleration have components $u_{\hat{0}} = -1$, $u_{\hat{j}} = 0$, $a_{\hat{0}} = 0$ in the observer's own proper frame, these connection coefficients are

$$\left.\begin{aligned} &\Gamma^{\hat{0}}{}_{\hat{0}\hat{0}} = \Gamma_{\hat{0}\hat{0}\hat{0}} = 0, \\ &\Gamma^{\hat{0}}{}_{\hat{j}\hat{0}} = -\Gamma_{\hat{0}\hat{j}\hat{0}} = +\Gamma_{\hat{j}\hat{0}\hat{0}} = +\Gamma^{\hat{j}}{}_{\hat{0}\hat{0}} = a^{\hat{j}}, \\ &\Gamma^{\hat{j}}{}_{\hat{k}\hat{0}} = \Gamma_{\hat{j}\hat{k}\hat{0}} = -\omega^{\hat{i}}\epsilon_{\hat{0}\hat{i}\hat{j}\hat{k}}, \end{aligned}\right\} \text{ all along } \mathscr{P}_0(\tau). \tag{13.69a}$$

The remaining connection coefficients can be read from the geodesic equation for the geodesics $\mathscr{G}[\tau, \boldsymbol{n}, s]$ that emanate from the observer's world line. According to equation (13.65), the coordinate representation of each such geodesic is

$$x^{\hat{0}}(s) = \tau = \text{constant}, \qquad x^{\hat{j}}(s) = n^{\hat{j}}s;$$

hence, $d^2x^{\hat{\alpha}}/ds^2 = 0$ all along the geodesic, and the geodesic equation reads

$$0 = \frac{d^2x^{\hat{\alpha}}}{ds^2} + \Gamma^{\hat{\alpha}}{}_{\hat{\beta}\hat{\gamma}} \frac{dx^{\hat{\beta}}}{ds}\frac{dx^{\hat{\gamma}}}{ds} = \Gamma^{\hat{\alpha}}{}_{\hat{j}\hat{k}} n^{\hat{j}} n^{\hat{k}}.$$

This equation is satisfied on the observer's world line for all spatial geodesics (all $n^{\hat{j}}$) if and only if

$$\Gamma^{\hat{\alpha}}{}_{\hat{j}\hat{k}} = \Gamma_{\hat{\alpha}\hat{j}\hat{k}} = 0 \text{ all along } \mathscr{P}_0(\tau). \tag{13.69b}$$

The values (13.69) of the connection coefficients determine uniquely the partial derivatives of the metric coefficients [see equation (13.19′)]:

$$\left.\begin{array}{ll} g_{\hat{\alpha}\hat{\beta},\hat{0}} = 0, & g_{\hat{j}\hat{k},\hat{l}} = 0, \\ g_{\hat{0}\hat{0},\hat{j}} = -2a_{\hat{j}}, & g_{\hat{0}\hat{j},\hat{k}} = -\epsilon_{\hat{0}\hat{j}\hat{k}\hat{l}}\omega^{\hat{l}} \end{array}\right\} \text{all along } \mathscr{P}_0(\tau); \tag{13.70}$$

Metric of proper reference frame, and its physical interpretation

and these, plus the orthonormality condition $g_{\hat{\alpha}\hat{\beta}}[\mathscr{P}_0(\tau)] = \eta_{\alpha\beta}$, imply that the line element near the observer's world line is

$$\begin{aligned} ds^2 = &-(1 + 2a_{\hat{j}}x^{\hat{j}})\, dx^{\hat{0}2} - 2(\epsilon_{\hat{j}\hat{k}\hat{l}}x^{\hat{k}}\omega^{\hat{l}})\, dx^{\hat{0}}\, dx^{\hat{j}} \\ &+ \delta_{\hat{j}\hat{k}}\, dx^{\hat{j}}\, dx^{\hat{k}} + O(|x^{\hat{j}}|^2)\, dx^{\hat{\alpha}}\, dx^{\hat{\beta}}. \end{aligned} \tag{13.71}$$

Several features of this line element deserve notice, as follows.

(1) On the observer's world line $\mathscr{P}_0(\tau)$—i.e., $x^{\hat{j}} = 0$—$ds^2 = \eta_{\alpha\beta}\, dx^{\hat{\alpha}}\, dx^{\hat{\beta}}$.

(2) The observer's acceleration shows up in a correction term to $g_{\hat{0}\hat{0}}$,

$$\delta g_{\hat{0}\hat{0}} = -2\boldsymbol{a}\cdot\boldsymbol{x}, \tag{13.72a}$$

which is proportional to distance along the acceleration direction. For the flat-spacetime derivation of this correction term, see §6.6.

(3) The observer's rotation relative to inertial-guidance gyroscopes shows up in a correction term to $g_{\hat{0}\hat{j}}$, which can be rewritten in 3-vector notation

$$\delta g_{\hat{0}\hat{j}}\boldsymbol{e}_{\hat{j}} = -\boldsymbol{x}\times\boldsymbol{\omega} = +\boldsymbol{\omega}\times\boldsymbol{x}. \tag{13.72b}$$

(4) These first-order corrections to the line element are unaffected by spacetime curvature and contain no information about curvature. Only at second order, $O(|x^{\hat{j}}|^2)$, will curvature begin to show up.

(5) *In the special case of zero acceleration and zero rotation* ($\boldsymbol{a} = \boldsymbol{\omega} = 0$), *the observer's proper reference frame reduces to a local Lorentz frame* ($g_{\hat{\alpha}\hat{\beta}} = \eta_{\alpha\beta}$, $\Gamma^{\hat{\alpha}}{}_{\hat{\beta}\hat{\gamma}} = 0$) *all along his geodesic world line!* By contrast, the local Lorentz coordinate

systems constructed earlier in the book ("general" local Lorentz coordinates of §8.6, "Riemann normal coordinates" of §11.6) are local Lorentz only at a single event.

In the case of zero rotation and zero acceleration, one can derive the following expression for the metric, accurate to second order in $|x^{\hat{j}}|$:

$$\begin{aligned} ds^2 = (-1 - R_{\hat{0}\hat{l}\hat{0}\hat{m}} x^{\hat{l}} x^{\hat{m}})\, dt^2 &- \left(\frac{4}{3} R_{\hat{0}\hat{l}\hat{j}\hat{m}} x^{\hat{l}} x^{\hat{m}}\right) dt\, dx^{\hat{j}} \\ &+ \left(\delta_{\hat{i}\hat{j}} - \frac{1}{3} R_{\hat{i}\hat{l}\hat{j}\hat{m}} x^{\hat{l}} x^{\hat{m}}\right) dx^{\hat{i}}\, dx^{\hat{j}} + O(|x^{\hat{j}}|^3)\, dx^{\hat{\alpha}}\, dx^{\hat{\beta}} \end{aligned} \tag{13.73}$$

[see, e.g., Manasse and Misner (1963)]. Here $R_{\hat{\alpha}\hat{\beta}\hat{\gamma}\hat{\delta}}$ are the components of the Riemann tensor along the world line $x^{\hat{j}} = 0$. Such coordinates are called "Fermi Normal Coordinates."

EXERCISES

Exercise 13.14. INERTIAL AND CORIOLIS FORCES

An accelerated observer studies the path of a freely falling particle as it passes through the origin of his proper reference frame. If

$$\boldsymbol{v} \equiv (dx^{\hat{j}}/dx^{\hat{0}})\boldsymbol{e}_{\hat{j}} \tag{13.74}$$

is the particle's ordinary velocity, show that its ordinary acceleration relative to the observer's proper reference frame is

$$\frac{d^2x^{\hat{j}}}{dx^{\hat{0}2}}\boldsymbol{e}_{\hat{j}} = -\underbrace{\boldsymbol{a}}_{\text{[inertial acceleration]}} - \underbrace{2\boldsymbol{\omega} \times \boldsymbol{v}}_{\text{[Coriolis acceleration]}} + \underbrace{2(\boldsymbol{a} \cdot \boldsymbol{v})\boldsymbol{v}}_{\text{[relativistic correction to inertial acceleration]}}. \tag{13.75}$$

Here $\boldsymbol{a}$ is the observer's own 4-acceleration, and $\boldsymbol{\omega}$ is the angular velocity with which his spatial basis vectors $\boldsymbol{e}_{\hat{j}}$ are rotating [see equations (13.62)]. [*Hint:* Use the geodesic equation at the point $x^{\hat{j}} = 0$ of the particle's trajectory. *Note:* This result was derived in flat spacetime in exercise 6.8 using a different method.]

Exercise 13.15. ROTATION GROUP: METRIC

(Continuation of exercises 9.13, 9.14, 10.17 and 11.12). Show that for the manifold $SO(3)$ of the rotation group, there exists a metric $\boldsymbol{g}$ that is compatible with the covariant derivative $\boldsymbol{\nabla}$. Prove existence by exhibiting the metric components explicitly in the noncoordinate basis of generators $\{\boldsymbol{e}_\alpha\}$. [*Answer:*

$$g_{\alpha\beta} = \delta_{\alpha\beta}. \tag{13.76}$$

Restated in words: If one postulates that: (1) the manifold of the rotation group is locally Euclidean; (2) the generators of infinitesimal rotations $\{\boldsymbol{e}_\alpha\}$ are orthonormal, $\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta = \delta_{\alpha\beta}$; and (3) $\{\boldsymbol{e}_\alpha\}$ obey the standard rotation-group commutation relations

$$[\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta] = -\epsilon_{\alpha\beta\gamma}\boldsymbol{e}_\gamma; \tag{13.77}$$

then the resulting geodesics of $SO(3)$ agree with the geodesics chosen in exercise 10.17.]

2

CHAPTER 14

CALCULATION OF CURVATURE

§14.1. CURVATURE AS A TOOL FOR UNDERSTANDING PHYSICS

Elementary physics sometimes allows one to shortcircuit any systematized calculation of curvature (frequency of oscillation of test particle; tide-producing acceleration near a center of attraction; curvature of a closed 3-sphere model universe; effect of parallel transport on gyroscope or vector; see Figures 1.1, 1.10, and 1.12, and Boxes 1.6 and 1.7); but on other occasions a calculation of curvature is the quickest way into the physics. This chapter is designed for such occasions. It describes three ways to calculate curvature and gives the components of the Einstein curvature tensor for a plane gravitational wave (Box 14.4, equation 5), for the Friedmann geometry of the universe (Box 14.5), and for Schwarzschild geometry, both static (exercise 14.13) and dynamic (exercise 14.16). These and other calculations of curvature elsewhere are indexed under "curvature tensors."

This chapter is entirely Track 2.
Chapter 4 (differential forms) and Chapter 10, 11, and 13 (differential geometry) are necessary preparation for §§14.5–14.6.
This chapter is needed as preparation for Chapter 15 (Bianchi identities).
It will be helpful in many applications of gravitation theory (Chapters 23–40).

It is enough to look at an expression for a 4-geometry as complicated as

$$\begin{aligned} ds^2 = &-(x/3^{1/2}L + y^2/12L^2)^{-3^{1/2}}\left(\int \frac{v\,dz}{z}\right)^{-1}(-z/L)^{3^{-1/2}}\,dt^2 \\ &+(x/3^{1/2}L + y^2/12L^2)^{1+3^{1/2}}\left(\int \frac{v\,dz}{z}\right)^{1+2/3^{1/2}}(-z/L)^{-1+3^{-1/2}}\,dx^2 \\ &+(x/3^{1/2}L + y^2/12L^2)^{2+3^{1/2}}\left(\int \frac{v\,dz}{z}\right)^{1+2/3^{1/2}}(-z/L)^{-3^{-1/2}}\,dy^2 \\ &+(x/3^{1/2}L + y^2/12L^2)^{3+3^{1/2}}\left(\int \frac{v\,dz}{z}\right)^{1+2/3^{1/2}}(-z/L)^{-2-3^{-1/2}}\times \\ &\times\left(\frac{v^2-1}{-1-z/L}\right)dz^2 \end{aligned} \quad (14.1)$$

Situations in which one must compute curvature

[Harrison (1959)] to realize that one might understand the physical situation better if one knew what the curvature is; similarly with any other complicated expressions for metrics that arise from solving Einstein's equations or that appear undigested in the literature. In any such case, the appropriate method often is: curvature first, understanding second.

Curvature is the simplest local measure of geometric properties (see Box 14.1). Curvature is therefore a good first step toward a more comprehensive picture of the spacetime in question.

One sometimes has an expression for a spacetime metric first, and then makes calculations of curvature to understand it. But more often one makes calculations of curvature, subject to specified conditions of symmetry in space and time, as an aid in arriving at an expression for a physically interesting metric (stars, Chapters 23 to 26; model cosmologies, Chapters 27 to 30; collapse and black holes, Chapters 31 to 34; and gravitational waves, Chapters 35 to 37).

"Standard procedure" for computing curvature

The basic "standard procedure for computing curvature" is illustrated in Box 14.2. Two formulas in Box 14.2, derived previously, are used in succession. The first (equations 1 and 2) has the form $\Gamma \sim g\,\partial g$ and provides the $\Gamma^{\mu}{}_{\alpha\beta}$. The other (equation 3) has the form $R \sim \partial\Gamma + \Gamma^2$ and gives the curvature components $R^{\mu}{}_{\nu\alpha\beta}$.

Methods of displaying curvature formulas

After the curvature components have been computed, there are helpful ways to present the results. (1) Form the Ricci tensor $R_{\mu\nu} = R^{\alpha}{}_{\mu\alpha\nu}$ and the scalar curvature $R = R^{\mu}{}_{\mu}$. (2) Form other invariants such as $R^{\mu\nu}{}_{\alpha\beta}R^{\alpha\beta}{}_{\mu\nu}$. (3) Form components $R^{\hat{\mu}\hat{\nu}}{}_{\hat{\alpha}\hat{\beta}}$ in a judiciously chosen orthonormal frame $\boldsymbol{\omega}^{\hat{\alpha}} = L^{\hat{\alpha}}{}_{\beta}\,\boldsymbol{d}x^{\beta}$, and (4) display $R^{[\hat{\mu}\hat{\nu}]}{}_{[\hat{\alpha}\hat{\beta}]}$ as a 6×6 matrix (in four dimensions; a 3×3 matrix in three dimensions) where $[\hat{\mu}\hat{\nu}] = [\hat{0}\hat{1}], [\hat{0}\hat{2}], [\hat{0}\hat{3}], [\hat{2}\hat{3}], [\hat{3}\hat{1}], [\hat{1}\hat{2}]$ labels the rows and $[\hat{\alpha}\hat{\beta}]$ labels the columns (exercises 14.14 and 14.15). (5) Last, but by far the most important for general relativity, form the Einstein tensor $G^{\hat{\mu}}{}_{\hat{\nu}}$ as described in §14.2.

The method of computation outlined above and described in more detail in Box 14.2 is used wherever it is quicker to employ a standard method than to learn or invent a better method. The standard method is always preferable for the student in a short course where physical insight has higher priority than technical facility.

Computation of curvature using a computer

It is, however, a dull method, better suited to computers than to people. Even the algebra can be handled by a computer (see Box 14.3).

EXERCISES

Exercise 14.1. CURVATURE OF A TWO-DIMENSIONAL HYPERBOLOID

Compute the curvature of the hyperboloid $t^2 - x^2 - y^2 = T^2 = \text{const}$ in 2 + 1 Minkowski spacetime with $ds_3{}^2 = -dt^2 + dx^2 + dy^2$. First show that intervals within this two-dimensional surface can be expressed in the form $ds^2 = T^2(d\alpha^2 + \sinh^2\alpha\, d\phi^2)$ by a suitable choice of coordinates α, ϕ, on the hyperboloid.

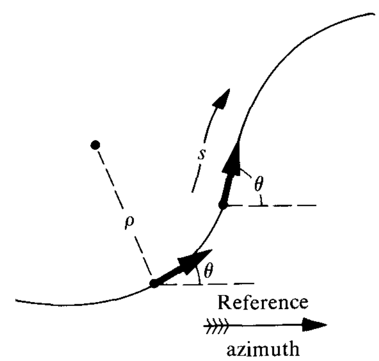
Exercise 14.2. RIEMANNIAN CURVATURE EXPRESSIBLE IN TERMS OF RICCI CURVATURE IN TWO AND THREE DIMENSIONS

In two dimensions, there is only one independent curvature component, R_{1212}. Evidently the single scalar quantity R must carry the same information. The two-dimensional identity $R_{\mu\nu\alpha\beta} = \frac{1}{2}R(g_{\mu\alpha}g_{\nu\beta} - g_{\mu\beta}g_{\nu\alpha})$ is established by noting that it is the only tensor formula giving

(*continued on page 343*)

Box 14.1 PERSPECTIVES ON CURVATURE

1. Historical point of departure: a curved line on a plane. There is no way to define the curvature of a line by measurements confined to ("intrinsic to") the line itself. One needs, for example, the azimuthal bearing θ of the tangent vector relative to a fixed direction in the plane, as a function of proper distance s measured along the curve; thus, $\theta = \theta(s)$. Then curvature κ and its reciprocal, the radius of curvature ρ, are given by $\kappa(s) = 1/\rho(s) = d\theta(s)/ds$. Alternatively, one can examine departure, y, measured normally off from the tangent line as a function of distance x measured along that tangent line; then $\kappa = 1/\rho = d^2y/dx^2$.



2. This concept was later extended to a curved surface embedded in flat (Euclidean) 3-space. Departure, z, of the smooth curved surface from the flat surface tangent to it at a given point is described in the neighborhood of that point by the quadratic expression

$$z = \frac{1}{2}ax^2 + bxy + \frac{1}{2}cy^2.$$

Rotation of the axes by an appropriate angle α,

$$x = \xi \cos\alpha + \eta \sin\alpha,$$
$$y = -\xi \sin\alpha + \eta \cos\alpha,$$

reduces this expression to

$$z = \frac{1}{2}\kappa_1\xi^2 + \frac{1}{2}\kappa_2\eta^2,$$

with

$$\kappa_1 = 1/\rho_1$$

and

$$\kappa_2 = 1/\rho_2$$

representing the two "principal curvatures" of the surface.

3. Gauss (1827) conceived the idea of defining curvature by measurements confined entirely to the surface ("society of ants"). From a given point $\mathscr{P}$ on the surface, proceed on a geodesic on the surface for a proper distance ϵ measured entirely within the surface. Repeat, starting at the original point but proceeding in other directions.

Box 14.1 ***(continued)***

Obtain an infinity of points. They define a "circle". Determine its proper circumference, again by measurements confined entirely to the surface. Using the metric corresponding to the embedding viewpoint

$$ds^2 = dz^2 + d\xi^2 + d\eta^2 \qquad \text{(Euclidean 3-space)}$$

$$= [(\kappa_1 \xi\, d\xi + \kappa_2 \eta\, d\eta)^2 + (d\xi^2 + d\eta^2)] \qquad \begin{pmatrix}\text{metric intrinsic}\\ \text{to the curved}\\ \text{2-geometry}\end{pmatrix},$$

one can calculate the result of such an "intrinsic measurement." One calculates that the circumference differs from the Euclidean value, $2\pi\epsilon$, by a fractional correction that is proportional to the square of ϵ; specifically,

$$\lim_{\epsilon \to 0} \frac{6}{\epsilon^2}\left(1 - \frac{\text{circumference}}{2\pi\epsilon}\right) = \kappa_1\kappa_2 = \frac{1}{\rho_1\rho_2} = \det\begin{pmatrix} a & b \\ b & c\end{pmatrix}.$$

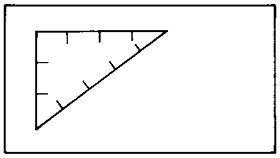
Note especially the first equality sign. Gauss did not conceal the elation he felt on discovering that something defined by measurements entirely within the surface agrees with the product of two quantities, κ_1 and κ_2, that individually demand for their definition measurements extrinsic to the surface.

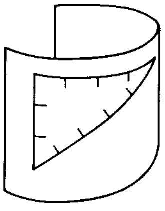
4. The contrast between "extrinsic" and "intrinsic" curvature is summarized in the terms,

$$\text{(extrinsic curvature)} = \kappa = (\kappa_1 + \kappa_2)(\text{cm}^{-1}),$$

$$\begin{pmatrix}\text{intrinsic or Gaussian}\\ \text{curvature}\end{pmatrix} = \kappa_1\kappa_2(\text{cm}^{-2})$$

(the latter being identical with half the scalar curvature invariant, R, of the 2-geometry). Draw a 3:4:5 triangle on a flat piece of paper; then curl up the paper. The Euclidean 2-geometry intrinsic to the piece of paper is preserved by this bending. The Gaussian curvature intrinsic to the surface remains unaltered; it keeps the Euclidean value of zero (κ_2, non-zero; κ_1, zero; product, $\kappa_1\kappa_2 =$ zero). However, the extrinsic curvature is changed from $\kappa_1 + \kappa_2 = 0$ to a non-zero value, $\kappa_1 + \kappa_2 \neq 0$.



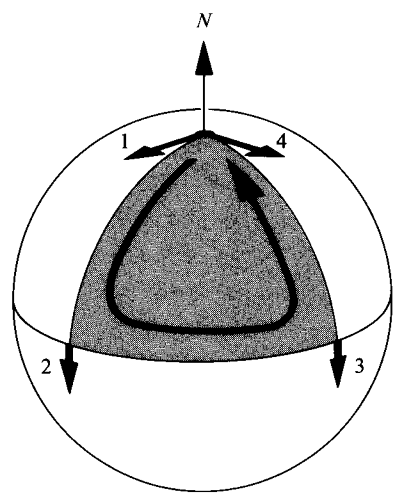


5. The curvature dealt with in this chapter is curvature intrinsic to spacetime; that is, curvature defined without any use of, and repelling every thought of, any embedding in any hypothetical higher-dimensional flat manifold (concept of Riemann,

Clifford, and Einstein that geometry is a dynamic participant in physics, not some God-given perfection above the battles of matter and energy).

6. The curvature of the geometry of spacetime imposes curvature on any spacelike slice (3-geometry; "initial-value hypersurface") through that spacetime (see "relations of Gauss and Codazzi" in Chapter 21, on the initial-value problem of geometrodynamics).

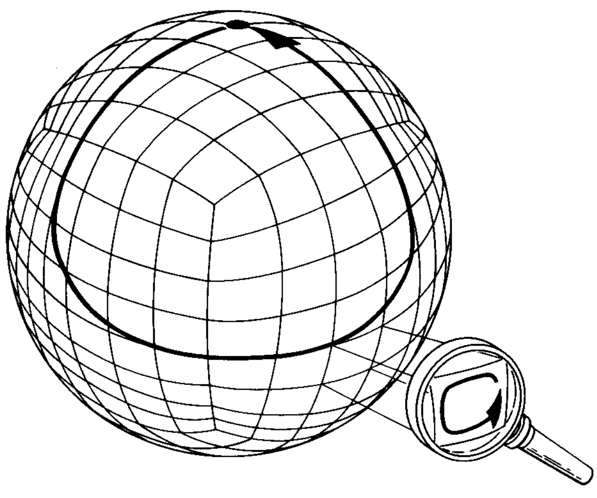
7. Rotation of a vector transported parallel to itself around a closed loop provides a definition of curvature as useful in four and three as in two dimensions. (In a curved two-dimensional geometry, at a point there is only one plane. Consequently only one number is required to describe the Gaussian curvature there. In three and four dimensions, there are more independent planes through a point and therefore more numbers are required to describe the curvature.) In the diagram, start with a vector at position 1 (North Pole). Transport it parallel to itself (position $2, 3, \ldots$) around a $90°$-$90°$-$90°$ spherical triangle. It arrives back at the starting point (position 4) turned through $90°$:



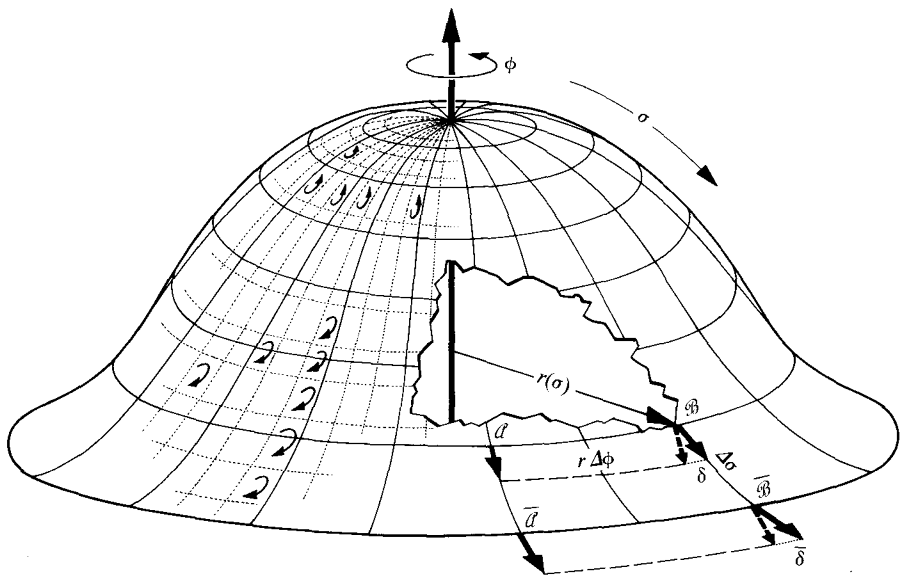
$$\begin{pmatrix}\text{Gaussian}\\ \text{curvature}\end{pmatrix} = \frac{\begin{pmatrix}\text{angle turned}\\ \text{through}\end{pmatrix}}{\begin{pmatrix}\text{area circum-}\\ \text{navigated}\end{pmatrix}} = \frac{(\pi/2)}{(1/8)(4\pi a^2)} = \frac{1}{a^2}$$

(positive; sense of rotation same as sense of circumnavigation).

8. Still staying for simplicity with a curved two-dimensional manifold, describe the curvature of the 2-surface as a 2-form ("box-like structure") defined over the entire surface. The number of boxes enclosed by any given route gives immediately the angle in radians (or tenths or hundredths of a radian, etc., depending on chosen fineness of subdivision) turned through by a vector carried parallel to itself around that route. The contribution of a given box is counted as positive or negative depending on whether the sense of the arrow marked on it (see magnified view) agrees or disagrees with the sense of circumnavigation of the route.



Box 14.1 (*continued*)



9. Curvature 2-form for the illustrated surface of rotational symmetry ("pith helmet") with metric $ds^2 = d\sigma^2 + r^2(\sigma)\, d\phi^2$ is

$$\textbf{\textit{curvature}} = -\frac{1}{r}\frac{d^2r}{d\sigma^2}\,\boldsymbol{d}\sigma \wedge r\,\boldsymbol{d}\phi \tag{1}$$

(positive on crown of helmet, negative around brim, as indicated by sense of arrows in the "boxes of the 2-form" shown at left). "Meaning" of r is illustrated by imbedding the surface in Euclidean 3-space, a convenience for visualization; but more important is the idea of a 2-geometry defined by measurements intrinsic to it, with no embedding.

10. How lengths ("metric") determine curvature in quantitative detail is shown nowhere more clearly than in this two-dimensional example, a model for "what is going on behind the scene" in the mathematical calculations done in this chapter with 1-forms and 2-forms in four-dimensional spacetime.

a. Net rotation in going around element of surface $\mathscr{A}\mathscr{B}\bar{\mathscr{B}}\bar{\mathscr{A}}\mathscr{A}$ is $\delta - \bar{\delta}$ (no turn of vector to left or to right in its transport along a meridian $\mathscr{A}\bar{\mathscr{A}}$ or $\mathscr{B}\bar{\mathscr{B}}$).

b. Rotation of vector in going from $\mathscr{A}$ to $\mathscr{B}$, relative to coordinate system (directions of meridians), is

$$(\text{angle } \delta) = \frac{\text{arc}}{\text{length}} = \frac{r(\sigma + d\sigma)\,\Delta\phi - r(\sigma)\,\Delta\phi}{d\sigma} = \left(\frac{dr}{d\sigma}\right)_{\sigma} \Delta\phi.$$

c. Rotation of vector in going from $\bar{\mathscr{A}}$ to $\bar{\mathscr{B}}$ is similarly

$$(\text{angle } \bar{\delta}) = \left(\frac{dr}{d\sigma}\right)_{\sigma + \Delta\sigma} \Delta\phi.$$

d. Thus net rotation is:

$$\delta - \bar{\delta} = -\left(\frac{d^2 r}{d\sigma^2}\right)_{\sigma} \Delta\sigma\, \Delta\phi.$$

e. Expressed as a form, this gives immediately equation (1).

f. Ideas and calculations are more complicated in four dimensions, primarily because one has to deal with different choices for the orientation of the surface to be studied at the point in question.

11. Translation of these geometric ideas into the language of forms is most immediate when one stays with this example of two dimensions. A sample vector $A^i = (A^1, A^2)$ carried around the boundary of an element of surface comes back to its starting point slightly changed in direction:

$$-\begin{pmatrix}\text{change}\\ \text{in } A^i\end{pmatrix} = \mathscr{R}^i{}_j A^j. \tag{2}$$

a. To be more specific, it is convenient to adopt as the basis 1-forms $\boldsymbol{\omega}^{\hat{1}} = \boldsymbol{d}\sigma$ and $\boldsymbol{\omega}^{\hat{2}} = r\,\boldsymbol{d}\phi$, and have $A^{\hat{1}}$ as the component of $\boldsymbol{A}$ along the direction of increasing σ, $A^{\hat{2}}$ as the component of $\boldsymbol{A}$ along the direction of increasing ϕ. The matrix $\mathscr{R}^{\hat{i}}{}_{\hat{j}}$ is a rotation matrix, which produces a change in direction but no change in length (zero diagonal components); thus here

$$\|\mathscr{R}^{\hat{i}}{}_{\hat{j}}\| = \left\|\begin{matrix} 0 & \mathscr{R}^{\hat{1}}{}_{\hat{2}} \\ -\mathscr{R}^{\hat{1}}{}_{\hat{2}} & 0 \end{matrix}\right\|. \tag{3}$$

In this example, $\mathscr{R}^{\hat{1}}{}_{\hat{2}}$ evidently represents the angle through which the vector $\boldsymbol{A}$ turns on transport parallel to itself around the element of surface.

b. So far the rotation is "indefinite" because the size of the element of surface has not yet been specified. It is most conveniently conceived as an elementary parallelogram, defined by two vectors ("bivector"). Thus $\mathscr{R}^{\hat{i}}{}_{\hat{j}}$, or, specifically, the one element that counts, $\mathscr{R}^{\hat{1}}{}_{\hat{2}}$ (the "angle of rotation"), has to be envisaged as a mathematical object ("2-form") endowed with two slots, into which these two vectors are inserted to get a definite number (angle in radians). In the example of the pith helmet, one has, from equation (1)

$$\mathscr{R}^{\hat{1}}{}_{\hat{2}} = -\frac{1}{r}\frac{d^2 r}{d\sigma^2}\,\boldsymbol{\omega}^{\hat{1}} \wedge \boldsymbol{\omega}^{\hat{2}}. \tag{4}$$

Thus the $\mathscr{R}^{\mu}{}_{\nu}$ in the text are called "curvature 2-forms."

Box 14.1 (*continued*)

c. The text tells one how to read out of such expressions the components of the Riemann curvature tensor; for example here,

$$R^{\hat{1}}{}_{\hat{2}\hat{1}\hat{2}} = -R^{\hat{1}}{}_{\hat{2}\hat{2}\hat{1}} = (-1/r)(d^2r/d\sigma^2) \text{ (coefficients of } \boldsymbol{\omega}^{\hat{1}} \wedge \boldsymbol{\omega}^{\hat{2}} \text{ or } \boldsymbol{\omega}^{\hat{2}} \wedge \boldsymbol{\omega}^{\hat{1}}).$$

d. Generalizing to four dimensions, one understands by $R^{\alpha}{}_{\beta\mu\nu}$ the factor that one has to multiply by three numbers to obtain a fourth. The number obtained is the change (with reversed sign) that takes place in the αth component of a vector when that vector is transported parallel to itself around a closed path, defined, for example, by a parallelogram built from two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. The factors that multiply $R^{\alpha}{}_{\beta\mu\nu}$ are (1) the component of the vector $\boldsymbol{A}$ in the βth direction and (2, 3) the $\mu\nu$ component of the extension of the parallelogram, $(u^{\mu}v^{\nu} - u^{\nu}v^{\mu})$. Thus

$$\delta A^{\alpha} = -R^{\alpha}{}_{\beta|\mu\nu|}A^{\beta}(u^{\mu}v^{\nu} - u^{\nu}v^{\mu}).$$

Box 14.2 STRAIGHTFORWARD CURVATURE COMPUTATION (Illustrated for a Globe)

The elementary and universally applicable method for computing the components $R^{\mu}{}_{\nu\alpha\beta}$ of the Riemann curvature tensor starts from the metric components $g_{\mu\nu}$ in a coordinate basis, and proceeds by the following scheme:

$$g_{\mu\nu} \xrightarrow{\Gamma\sim\partial g} \Gamma_{\mu\alpha\beta} \longrightarrow \Gamma^{\mu}{}_{\alpha\beta} \xrightarrow{R\sim\partial\Gamma+\Gamma\Gamma} R^{\mu}{}_{\nu\alpha\beta}.$$

The formulas required for these three steps are

$$\Gamma_{\mu\alpha\beta} = \frac{1}{2}\left(\frac{\partial g_{\mu\alpha}}{\partial x^{\beta}} + \frac{\partial g_{\mu\beta}}{\partial x^{\alpha}} - \frac{\partial g_{\alpha\beta}}{\partial x^{\mu}}\right), \tag{1}$$

$$\Gamma^{\mu}{}_{\alpha\beta} = g^{\mu\nu}\Gamma_{\nu\alpha\beta}, \tag{2}$$

and

$$R^{\mu}{}_{\nu\alpha\beta} = \frac{\partial\Gamma^{\mu}{}_{\nu\beta}}{\partial x^{\alpha}} - \frac{\partial\Gamma^{\mu}{}_{\nu\alpha}}{\partial x^{\beta}} + \Gamma^{\mu}{}_{\rho\alpha}\Gamma^{\rho}{}_{\nu\beta} - \Gamma^{\mu}{}_{\rho\beta}\Gamma^{\rho}{}_{\nu\alpha}. \tag{3}$$

The metric of the two-dimensional surface of a sphere of radius a is

$$ds^2 = a^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{4}$$

To compute the curvature by the standard method, use the formula for ds^2 as a table of g_{kl} values. It shows that $g_{\theta\theta} = a^2$, $g_{\theta\phi} = 0$, $g_{\phi\phi} = a^2\sin^2\theta$. Compute the six possible different $\Gamma_{jkl} = \Gamma_{jlk}$ (there will be 40 in four dimensions) from formula

(1). Thus

$$\begin{aligned} \Gamma_{\theta\phi\phi} &= -a^2 \sin\theta\cos\theta = -\Gamma_{\phi\phi\theta}, \\ \Gamma_{\theta\theta\theta} &= \Gamma_{\phi\phi\phi} = 0, \\ \Gamma_{\theta\theta\phi} &= \Gamma_{\phi\theta\theta} = 0. \end{aligned} \tag{5}$$

Raise the first index:

$$\begin{aligned} \Gamma^{\theta}{}_{\phi\phi} &= -\sin\theta\cos\theta, \\ \Gamma^{\phi}{}_{\phi\theta} &= \cot\theta, \\ \Gamma^{\theta}{}_{\theta\theta} &= \Gamma^{\theta}{}_{\theta\phi} = 0 = \Gamma^{\phi}{}_{\theta\theta} = \Gamma^{\phi}{}_{\phi\phi}. \end{aligned} \tag{6}$$

Choose a suitable curvature component (one that is not automatically zero by reason of the elementary symmetry $R_{\mu\nu\alpha\beta} = R_{[\mu\nu][\alpha\beta]}$, nor previously computed in another form, as by $R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu}$). In this two-dimensional example, there is only one choice (compared to 21 such computations in four dimensions); it is

$$\begin{aligned} R^{\theta}{}_{\phi\theta\phi} &= \frac{\partial\Gamma^{\theta}{}_{\phi\phi}}{\partial\theta} - \frac{\partial\Gamma^{\theta}{}_{\phi\theta}}{\partial\phi} + \Gamma^{\theta}{}_{k\theta}\Gamma^{k}{}_{\phi\phi} - \Gamma^{\theta}{}_{k\phi}\Gamma^{k}{}_{\phi\theta} \\ &= \frac{\partial\Gamma^{\theta}{}_{\phi\phi}}{\partial\theta} - 0 + 0 - \Gamma^{\theta}{}_{\phi\phi}\Gamma^{\phi}{}_{\phi\theta} \\ &= \sin^2\theta - \cos^2\theta + \sin\theta\cos\theta\cot\theta; \end{aligned}$$

so

$$R^{\theta}{}_{\phi\theta\phi} = \sin^2\theta \tag{7}$$

or

$$R^{\theta\phi}{}_{\theta\phi} = \frac{1}{a^2}. \tag{8}$$

Contraction gives the components of the Ricci tensor,

$$R^{\theta}{}_{\theta} = R^{\phi}{}_{\phi} = \frac{1}{a^2}, \qquad R^{\theta}{}_{\phi} = 0, \tag{9}$$

and further contraction gives the curvature scalar

$$R = 2/a^2. \tag{10}$$

A convenient orthonormal frame in this manifold is

$$\boldsymbol{\omega}^{\hat\theta} = a\,\boldsymbol{d}\theta, \qquad \boldsymbol{\omega}^{\hat\phi} = a\sin\theta\,\boldsymbol{d}\phi. \tag{11}$$

More generally one writes $\boldsymbol{\omega}^{\hat\alpha} = L^{\hat\alpha}{}_{\beta}\,\boldsymbol{d}x^{\beta}$. To transform the curvature tensor to orthonormal components in this simple but illuminating example of a diagonal metric requires a single normalization factor for each index on a tensor. Thus $v^{\hat\theta} = av^{\theta}$, $v^{\hat\phi} = a\sin\theta\, v^{\phi}$, $v_{\hat\theta} = a^{-1}v_{\theta}$, $v_{\hat\phi} = (a\sin\theta)^{-1}v_{\phi}$. Similarly, from $R^{\theta}{}_{\phi\theta\phi} = \sin^2\theta$ one finds the components of the curvature tensor,

$$R^{\hat\theta}{}_{\hat\phi\hat\theta\hat\phi} = \frac{1}{a^2} = R^{\hat\theta\hat\phi}{}_{\hat\theta\hat\phi}, \tag{12}$$

in the orthonormal frame.

Box 14.3 ANALYTICAL CALCULATIONS ON A COMPUTER

Research in gravitation physics and general relativity is sometimes beset by long calculations, requiring meticulous care, of such quantities as the Einstein and Riemann curvature tensors for a given metric, or the divergence of a given stress-energy tensor, or the Newman-Penrose tetrad equations under given algebraic assumptions. Such calculations are sufficiently straightforward and deductive in logical structure that they can be handled by a computer. Since 1966, computers have been generally taking over such tasks.

There are several computer languages in which the investigator can program his analytic calculations. The computer expert may wish to work in a machine-oriented language such as LISP [see, e.g., the work of Fletcher (1966) and of Hearn (1970)]. However, most appliers of relativity will prefer user-oriented languages such as REDUCE [created by Hearn (1970) and available for the IBM 360 and 370, and the PDP 10 computers], ALAM [created by D'Inverno (1969) and available on Atlas computers], CAMAL [created by Barton, Bourne, and Fitch (1970) and available on Atlas computers], and FORMAC [created by Tobey *et al.* (1967) and available on IBM 7090, 7094, 360, and 370]. For a review of activity in this area, see Barton and Fitch (1971). Here we describe only FORMAC. It is the most widely available and widely used of the languages; but it is probably *not* the most powerful [see, e.g., D'Inverno (1969)]. FORMAC is to analytic work what the earliest and most primitive versions of FORTRAN were to numerical work.

FORMAC manipulates algebraic expressions involving: numerical constants, such as 1/3; symbolic constants, such as x or u; specific elementary functions, such as $\sin(u)$ or $\exp(x)$; and symbolic functions of several variables, such as $f(x, u)$ or $g(u)$. For example, it can add $ax + bx^2$ to $2x + (3 + b)x^2$ and get $(a + 2)x + (3 + 2b)x^2$; it can take the partial derivative of $x^2uf(x, u) + \cos(x)$ with respect to x and get

$$2xuf(x, u) + x^2u\, \partial f(x, u)/\partial x - \sin(x).$$

It can do any algebraic or differential-calculus computation that a human can do—but without making mistakes! Unfortunately, it cannot integrate analytically; integration requires inductive logic rather than deductive logic.

PL/1 is a language that can be used simultaneously with FORMAC or independently of it. PL/1 manipulates strings of characters—e.g., "Z/1×29−+/." It knows symbolic logic; it can tell whether two strings are identical; it can insert new characters into a string or remove old ones; but it does not know the rules of algebra or differential calculus. Thus, its primary use is as an adjunct to FORMAC (though from the viewpoint of the computer system FORMAC is an adjunct of PL/1).

FORMAC programs for evaluating Einstein's tensor in terms of given metric components and for doing other calculations are available from many past users [see, e.g., Fletcher, Clemens, Matzner, Thorne, and Zimmerman (1967); Ernst (1968); Harrison (1970)]. However, programming in FORMAC is sufficiently simple that one ordinarily does not have difficulty creating one's own program to do a given task. If a difficulty does arise, it may be because the analytic computation exhausts the core of the computer. It is easy to create an expression too large to fit in the core of any existing computer by several differentiations of an expression half a page long!

Users of FORMAC, confronted by core-exhaustion, have devised several ways to solve their problems. One is to remove unneeded parts of the program and of the FORMAC system from the core. Routines called PURGE and KILL have been developed for this purpose by Clemens and Matzner (1967). Another is to create the answer to a given calculation in manageable-sized pieces and output those pieces from the computer's core onto its disk. One must then add all the pieces together—a task that is impossible using FORMAC alone, or even FORMAC plus PL/1, but a task that James Hartle has solved [see Hartle and Thorne (1974)] by using a combination of FORMAC, PL/1, and IBM data-manipulation routines called SORT.

$R_{\mu\nu\alpha\beta}$ as a linear function of R, constructed from R and the metric alone, and with the correct contracted value $R^{\mu\nu}{}_{\mu\nu} = R$. Establish a corresponding three-dimensional identity expressing $R_{ijk\ell}$ in terms of the Ricci tensor R_{jk} and the metric.

Exercise 14.3. CURVATURE OF 3-SPHERE IN ORTHONORMAL FRAME

Compute the curvature tensor for a 3-sphere

$$ds^2 = a^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)] \tag{14.2}$$

or for a 3-hyperboloid

$$ds^2 = a^2[d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{14.3}$$

Convert the coordinate-based components $R^i{}_{jk\ell}$ to a corresponding orthonormal basis, $R^{\hat{i}}{}_{\hat{j}\hat{k}\hat{\ell}}$. Display $R^{\hat{i}\hat{j}}{}_{\hat{k}\hat{\ell}} = R^{[\hat{i}\hat{j}]}{}_{[\hat{k}\hat{\ell}]}$ as a 3×3 matrix with appropriately labeled rows and columns.

§14.2. FORMING THE EINSTEIN TENSOR

The distribution of matter in space does not immediately tell all details of the local curvature of space, according to Einstein. The stress-energy tensor provides information only about a certain combination of components of the Riemann curvature tensor, the combination that makes up the Einstein tensor. Chapter 13 described two equivalent ways to calculate the Einstein tensor: (1) by successive contractions of the Riemann tensor

Three ways to compute the Einstein tensor from the Riemann tensor

$$\begin{aligned} R_{\mu\nu} &= R^\alpha{}_{\mu\alpha\nu}, \qquad R = g^{\mu\nu}R_{\mu\nu}, \\ G_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \end{aligned} \tag{14.4}$$

[equations (13.48) and (13.49)]; (2) by forming the dual of the Riemann tensor and then contracting:

$$\begin{aligned} \mathcal{G}_{\alpha\beta}{}^{\gamma\delta} \equiv ({}^*R^*)_{\alpha\beta}{}^{\gamma\delta} &= \epsilon_{\alpha\beta\mu\nu} R^{|\mu\nu|}{}_{|\rho\sigma|}\epsilon^{\rho\sigma\gamma\delta} \\ &= -\delta^{\rho\sigma\gamma\delta}{}_{\alpha\beta\mu\nu} R^{|\mu\nu|}{}_{|\rho\sigma|}, \end{aligned} \tag{14.5a}$$

$$G_\beta{}^\delta = \mathcal{G}_{\alpha\beta}{}^{\alpha\delta} \tag{14.5b}$$

[equations (13.46) and (13.47)]. A third method, usually superior to either of these, is discovered by combining equations (14.5a,b):

$$G_\beta{}^\delta = G^\delta{}_\beta = -\delta^{\delta\rho\sigma}{}_{\beta\mu\nu} R^{|\mu\nu|}{}_{|\rho\sigma|}. \tag{14.6}$$

[Note: in any frame, orthonormal or not, the permutation tensor $\delta^{\delta\rho\sigma}{}_{\beta\mu\nu}$ has components

$$\delta^{\delta\rho\sigma}{}_{\beta\mu\nu} = \delta_{\beta\mu\nu}{}^{\delta\rho\sigma} = \begin{cases} +1 \text{ if } \delta\rho\sigma \text{ is an even permutation of } \beta\mu\nu, \\ -1 \text{ if } \delta\rho\sigma \text{ is an odd permutation of } \beta\mu\nu, \\ 0 \text{ otherwise;} \end{cases}$$

to see this, simply evaluate $\delta^{\delta\rho\sigma}{}_{\beta\mu\nu}$ using definition (3.50h) and using the components (8.10) of $\epsilon_{\alpha\beta\mu\nu}$ and $\epsilon^{\rho\sigma\gamma\delta}$.] Equation (14.6) for the Einstein tensor, written out explicitly, reads

$$\begin{aligned} G^0{}_0 &= -(R^{12}{}_{12} + R^{23}{}_{23} + R^{31}{}_{31}), \\ G^1{}_1 &= -(R^{02}{}_{02} + R^{03}{}_{03} + R^{23}{}_{23}), \\ G^0{}_1 &= R^{02}{}_{12} + R^{03}{}_{13}, \\ G^1{}_2 &= R^{10}{}_{20} + R^{13}{}_{23}, \end{aligned} \tag{14.7}$$

and every other component is given by a similar formula, obtainable by obvious permutations of indices.

§14.3. MORE EFFICIENT COMPUTATION

Standard method of computing curvature is wasteful

If the answer to a problem or the result of a computation is not simple, then there is no simple way to obtain it. But when a long computation gives a short answer, *then* one looks for a better method. Many of the best-known applications of general relativity present one with metric forms in which many of the components $g_{\mu\nu}$, $\Gamma^\mu{}_{\alpha\beta}$, and $R^\mu{}_{\nu\alpha\beta}$ are zero; for them the standard computation of the curvature (Box 14.2) involves much "wasted" effort. One computes many $\Gamma^\mu{}_{\alpha\beta}$ that turn out to be zero. One checks off many terms in a sum like $-\Gamma^\mu{}_{\rho\beta}\Gamma^\rho{}_{\alpha\mu}$ that are zero, or cancel with others to give zero. Two alternative procedures are available to eliminate some of this "waste." The "geodesic Lagrangian" method provides an economical way to tabulate the $\Gamma^\mu{}_{\alpha\beta}$. The method of "curvature 2-forms" reorganizes the description from beginning to end, and computes both the connection and the curvature.

Ways to avoid "waste":

(1) geodesic Lagrangian method

(2) method of curvature 2-forms

The geodesic Lagrangian method is only a moderate improvement over the standard method, but it also demands only a modest investment in the calculus of variations, an investment that pays off in any case in other contexts in the world of mathematics and physics. In contrast, the method of curvature 2-forms is efficient, but demands a heavier investment in the mathematics of 1-forms and 2-forms than anyone would normally find needful for any introductory survey of relativity. Anyone facing several days' work at computing curvatures, however, would do well to learn the algorithm of the curvature 2-forms.

§14.4. THE GEODESIC LAGRANGIAN METHOD

One normally thinks that the connection coefficients $\Gamma^\mu{}_{\alpha\beta}$ must be known before one can write the geodesic equation

$$\ddot{x}^\mu + \Gamma^\mu{}_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = 0. \tag{14.8}$$

(Here and below dots denote derivative with respect to the affine parameter, λ.) However, the argument can be reversed. Once the geodesic equations have been written down, the connection coefficients can be read out of them. For instance, on the 2-sphere as treated in Box 14.2, the geodesic equations are

$$\ddot{\theta} - \sin\theta\cos\theta\,\dot{\phi}^2 = 0, \tag{14.9θ}$$

$$\ddot{\phi} + 2\cot\theta\,\dot{\phi}\dot{\theta} = 0. \tag{14.9ϕ}$$

The first equation here shows that $\Gamma^{\theta}{}_{\phi\phi} = -\sin\theta\cos\theta$; the second equation shows that $\Gamma^{\phi}{}_{\phi\theta} = \Gamma^{\phi}{}_{\theta\phi} = \cot\theta$; and the absence of any further terms shows that all other $\Gamma^{i}{}_{jk}$ are zero.

The first essential principle is thus clear: an explicit writing out of the geodesic equation is equivalent to a tabulation of all the connection coefficients $\Gamma^{\mu}{}_{\alpha\beta}$.

The second principle says more: one can write out the geodesic equation without ever having computed the $\Gamma^{\mu}{}_{\alpha\beta}$. In order to arrive at the equations for a geodesic (see Box 13.3), one need only recall that a geodesic is a parametrized curve that extremizes the integral

$$I = \frac{1}{2}\int g_{\mu\nu}\dot{x}^{\mu}\dot{x}^{\nu}\,d\lambda \tag{14.10}$$

in the sense

$$\delta I = 0.$$

Geodesic Lagrangian method in 4 steps:

In practical applications of this variational principle, *the first step is to rewrite equation (14.10) in the simplest possible form, inserting the specific values of $g_{\mu\nu}$ for the problem at hand.* If one's interest attaches to the geodesics themselves, one can recognize many constants of motion even without carrying out any variations (see Chapter 25 on geodesic motion in Schwarzschild geometry, especially §25.2 on conservation laws and constants of motion). For the purpose of computing the $\Gamma^{\mu}{}_{\alpha\beta}$, one proceeds to vary each coordinate in turn, obtaining four equations. Next these equations are rearranged so that their leading terms are $\ddot{x}^{\mu}$. In this form they must be precisely the geodesic equations (14.8). Consequently, the $\Gamma^{\mu}{}_{\alpha\beta}$ are immediately available as the coefficients in these four equations. For the final step in computing curvature by this method, one returns to the standard method and to formulas of the type $R \sim \partial\Gamma + \Gamma\Gamma$, treated in the standard way (Box 14.2); and as the need arises for each Γ in turn, one scans the geodesic equation to find it. The procedure is best understood by following an example: Box 14.4 provides one.

(1) write I in simple form

(2) vary I to get geodesic equation

(3) read off $\Gamma^{\alpha}{}_{\beta\gamma}$

(4) compute $R^{\alpha}{}_{\beta\gamma\delta}$ etc. by standard method

---

EXERCISE

Exercise 14.4. EINSTEIN EQUATIONS FOR THE CLOSED FRIEDMANN UNIVERSE CALCULATED BY USING THE GEODESIC LAGRANGIAN METHOD

The line element of interest here is (see Chapter 27)

$$ds^2 = -dt^2 + a^2(t)\,[d\chi^2 + \sin^2\chi\,(d\theta^2 + \sin^2\theta\,d\phi^2)].$$

(continued on page 348)

Box 14.4 GEODESIC LAGRANGIAN METHOD SHORTENS SOME CURVATURE COMPUTATIONS

Aim: Compute the curvature for the line element

$$ds^2 = L^2(e^{2\beta}\,dx^2 + e^{-2\beta}\,dy^2) - 2\,du\,dv \tag{1}$$

where L and β are functions of u only. [This metric is discussed as an example of a gravitational wave in §§35.9–35.12.]

Method: Obtain the $\Gamma^{\mu}{}_{\alpha\beta}$ from the geodesic equations as inferred from the variational principle (14.10), then compute $R^{\mu}{}_{\nu\alpha\beta} \sim \partial\Gamma + \Gamma^2$ as in Box 14.2.

Step 1. State the variational integral. For the metric under consideration, equation (14.10) requires $\delta I = 0$ for

$$I = \int\left[\frac{1}{2}L^2(e^{2\beta}\dot{x}^2 + e^{-2\beta}\dot{y}^2) - \dot{u}\dot{v}\right]d\lambda. \tag{2}$$

A world line that extremizes this integral is a geodesic.

Step 2: Vary the coordinates of the world line, one at a time, in their dependence on λ. First vary $x(\lambda)$, keeping fixed the functions $y(\lambda)$, $u(\lambda)$, and $v(\lambda)$. Then

$$\delta I = \int (L^2e^{2\beta}\dot{x})\,\delta\dot{x}\,d\lambda = -\int (L^2e^{2\beta}\dot{x})^{\cdot}\;\delta x\,d\lambda.$$

The requirement that $\delta I = 0$ for this variation (among others) gives

$$0 = (L^2e^{2\beta}\dot{x})^{\cdot} = L^2e^{2\beta}\ddot{x} + \dot{x}\dot{u}\frac{\partial}{\partial u}(L^2e^{2\beta}).$$

Varying y, u, v, in the same way gives

$$0 = (L^2e^{-2\beta}\dot{y})^{\cdot} = L^2e^{-2\beta}\ddot{y} + \dot{y}\dot{u}\frac{\partial}{\partial u}(L^2e^{-2\beta}),$$

$$0 = \ddot{v} + \frac{1}{2}\dot{x}^2\frac{\partial}{\partial u}(L^2e^{2\beta}) + \frac{1}{2}\dot{y}^2\frac{\partial}{\partial u}(L^2e^{-2\beta}),$$

$$0 = \ddot{u}.$$

Step 3: Rearrange to get $\ddot{x}^{\mu}$ leading terms. If this step is not straightforward, this method will not save time, and the technique of either Box 14.2 or Box 14.5 will be more suitable. In the example here, one quickly writes, using a prime for $\partial/\partial u$,

$$0 = \ddot{x} + 2(L^{-1}L' + \beta')\dot{x}\dot{u}, \tag{3x}$$

$$0 = \ddot{y} + 2(L^{-1}L' - \beta')\dot{y}\dot{u}, \tag{3y}$$

$$0 = \ddot{v} + (L^2e^{2\beta})(L^{-1}L' + \beta')\dot{x}^2 + (L^2e^{-2\beta})(L^{-1}L' - \beta')\dot{y}^2, \tag{3v}$$

$$0 = \ddot{u}. \tag{3u}$$

Step 3′: Interpret these equations as a tabulation of $\Gamma^{\mu}{}_{\alpha\beta}$. Equations (3) are the standard equations for a geodesic,

$$\ddot{x}^{\mu} + \Gamma^{\mu}{}_{\alpha\beta}\dot{x}^{\alpha}\dot{x}^{\beta} = 0.$$

Therefore it is enough to scan them to find the value of any desired Γ. For instance $\Gamma^{x}{}_{yu}$ must appear in the coefficient $(\Gamma^{x}{}_{yu} + \Gamma^{x}{}_{uy}) = 2\Gamma^{x}{}_{yu}$ of the $\dot{y}\dot{u}$ term in the equation for $\ddot{x}$. But no $\dot{y}\dot{u}$ term appears in equation ($3x$). Therefore $\Gamma^{x}{}_{yu}$ is zero in this example. Note that equations (3) are simple, in the sense that they contain few terms; therefore most of the $\Gamma^{\mu}{}_{\alpha\beta}$ must be zero. For instance, it follows from equation ($3u$) that all ten $\Gamma^{u}{}_{\alpha\beta}$ are zero. The only non-zero Γ's are $\Gamma^{x}{}_{xu} = \Gamma^{x}{}_{ux} = (L^{-1}L' + \beta')$ from equation ($3x$), $\Gamma^{y}{}_{yu} = \Gamma^{y}{}_{uy} = (L^{-1}L' - \beta')$ from equation ($3y$), and $\Gamma^{v}{}_{xx}$ and $\Gamma^{v}{}_{yy}$ from equation ($3v$).

Step 4: Compute each $R^{\mu}{}_{\nu\alpha\beta}$, etc. There is little relief from routine in systematically applying equation (3) from Box 14.2. One must list 21 components $R^{\mu}{}_{\nu\alpha\beta}$ that are not related by any of the symmetries $R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu} = -R_{\mu\nu\beta\alpha}$, and compute each. In the example here, one notes that $\Gamma^{u}{}_{\alpha\beta} = 0$ implies $R^{u}{}_{\alpha\beta\gamma} = -R_{v\alpha\beta\gamma} = 0$. Therefore 15 of the list of 21 vanish at one swat. The list then is:

$$\begin{aligned}
R_{v\alpha\beta\gamma} &= -R^{u}{}_{\alpha\beta\gamma} = 0,\\
R_{uxux} &= -R^{v}{}_{xux} = -(\Gamma^{v}{}_{xx})' + \Gamma^{v}{}_{xx}\Gamma^{x}{}_{xu}\\
&\qquad = -(L^2e^{2\beta})\left(\frac{L''}{L} + \beta'' + 2\frac{L'}{L}\beta' + \beta'^2\right),\\
R_{uxxy} &= -R^{v}{}_{xxy} = 0,\\
R_{uxyu} &= -R^{v}{}_{xyu} = 0,\\
R_{uyuy} &= -R^{v}{}_{yuy} = -(\Gamma^{v}{}_{yy})' + \Gamma^{v}{}_{yy}\Gamma^{y}{}_{yu}\\
&\qquad = -(L^2e^{-2\beta})\left(\frac{L''}{L} - \beta'' - 2\frac{L'}{L}\beta' + \beta'^2\right),\\
R_{uyxy} &= -R^{v}{}_{yxy} = 0,\\
R_{xyxy} &= (L^2e^{2\beta})R^{x}{}_{yxy} = 0.
\end{aligned} \tag{4}$$

One can now calculate the Einstein tensor via equation (14.7). In the example here, however, it is equally simple to form first the Ricci tensor by the straightforward contraction $R^{\mu}{}_{\alpha\mu\beta}$. Only $\mu = x$ and $\mu = y$ give any contribution, because no superscript index can be a u, and no subscript a v. Thus one finds

$$\begin{aligned} R_{uu} &= -2[L^{-1}L'' + \beta'^2],\\ &\text{all other } R_{\alpha\beta} = 0, \end{aligned} \tag{5}$$

and

$$R = 0. \tag{6}$$

From this last result, it follows that here the desired Einstein tensor is identical with the Ricci tensor.

(a) Set up the variational integral (14.10) for a geodesic in this metric, then successively vary t, χ, θ, and ϕ to obtain, after some rearrangement, four equations $0 = \ddot{t} + \cdots$, $0 = \ddot{\chi} + \cdots$, etc. displaying the Γ's in the form of equation (14.8).

(b) Use this display as a table of Γ's to compute $R^{t}{}_{\chi\mu\nu}$ and $R^{\chi}{}_{\theta\mu\nu}$, of which only $R^{t}{}_{\chi t\chi}$ and $R^{\chi}{}_{\theta\chi\theta}$ are non-zero (consequence of the complete equivalence of all directions tangent to the $\chi\theta\phi$ sphere).

(c) Convert to an orthonormal frame with $\boldsymbol{\omega}^{\hat{t}} = \boldsymbol{d}t$, $\boldsymbol{\omega}^{\hat{\chi}} = a\,\boldsymbol{d}\chi$, $\boldsymbol{\omega}^{\hat{\theta}} = ?$, $\boldsymbol{\omega}^{\hat{\phi}} = ?$, and list $R^{\hat{t}\hat{\chi}}{}_{\hat{t}\hat{\chi}}$ and $R^{\hat{\chi}\hat{\theta}}{}_{\hat{\chi}\hat{\theta}}$. Explain why all other components are known by symmetry in terms of these two.

(d) Calculate, using equations (14.7), all independent components of the Einstein tensor $G^{\hat{\mu}}{}_{\hat{\nu}}$. [*Answer:* See Box 14.5.]

---

§14.5. CURVATURE 2-FORMS

In electrodynamics the abstract notation

$$\boldsymbol{F} = \boldsymbol{d}\boldsymbol{A}$$

saves space compared to the explicit notation

$$F_{31} = \frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1},$$

$$F_{12} = \frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2},$$

$$\ldots, \text{etc. (six equations)};$$

Concepts needed for method of curvature 2-forms

there is no reason to shun similar economies in dealing with the dynamics of geometry. Cartan introduced the decisive ideas, seen above, of differential forms (where a simple object replaces a listing of four components; thus, $\boldsymbol{\sigma} = \sigma_\mu\, \boldsymbol{d}x^\mu$), and of the exterior derivative $\boldsymbol{d}$. He went on (1928, 1946) to package the 21 components $R_{\mu\nu\alpha\beta}$ of the curvature tensor into six curvature 2-forms,

$$\mathcal{R}^{\mu\nu} = -\mathcal{R}^{\nu\mu}.$$

Regarded purely as notation, these 2-forms automatically produce a profit. They cut down the weight of paper work required to list one's answer after one has it. They also provide a route into deeper insight on "curvature as a geometric object," although that is not the objective of immediate concern in this chapter.

Cartan's exterior derivative $\boldsymbol{d}$ automatically effects many cancelations in the calculation of curvature. It often cancels terms before they ever need to be evaluated.

Extension of Cartan's calculus from electromagnetism and other applications (Chapter 4) to the analysis of curvature (this chapter) requires two minor additions to the armament of forms and exterior derivative: (1) the idea of a vector-valued (or tensor-valued) exterior differential form; and (2) a corresponding generalization of the exterior derivative $\boldsymbol{d}$. This section uses both these tools in deriving the key formulas (14.18), (14.25), (14.31), and (14.32). Once derived, however, these formulas demand no more than the standard exterior derivative for all applications and for all calculations of curvature (§14.6 and Box 14.5).

Extended exterior derivative:

The extended exterior derivative leads to nothing new in the first two contexts to which one applies it: a scalar function ("0-form") and a vector field ("vector-valued 0-form"). Thus, take any function f. Its derivative in an unspecified direction is a 1-form; or, to make a new distinction that will soon become meaningful, a "scalar-valued 1-form." Specify the direction in which differentiation is to occur ("fill in the slot in the 1-form"). Thereby obtain the ordinary derivative as it applies to a function

(1) acting on a scalar

$$\langle \boldsymbol{d}f, \boldsymbol{u}\rangle = \partial_{\boldsymbol{u}} f. \tag{14.11}$$

Next, take any vector field $\boldsymbol{v}$. Its covariant derivative in an unspecified direction is a "vector-valued 1-form." Specify the direction $\boldsymbol{u}$ in which differentiation is to occur ("fill in the slot in the 1-form"). Thereby obtain the covariant derivative

(2) acting on a vector

$$\langle \boldsymbol{d}\boldsymbol{v}, \boldsymbol{u}\rangle \equiv \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}. \tag{14.12a}$$

This object too is not new; it is the covariant derivative of the vector $\boldsymbol{v}$ taken in the direction of the vector $\boldsymbol{u}$. When one abstracts away from any special choice of the direction of differentiation $\boldsymbol{u}$, one finds an expression that one has encountered before, though not under its new name of "vector-valued 1-form." This expression measures the covariant derivative of the vector $\boldsymbol{v}$ in an unspecified direction ("slot for direction not yet filled in"). From a look at (14.12a), one sees that this extended exterior derivative is applied to $\boldsymbol{v}$, without reference to $\boldsymbol{u}$, is

$$\boldsymbol{d}\boldsymbol{v} = \boldsymbol{\nabla}\boldsymbol{v}. \tag{14.12b}$$

Similarly, for any "tensor-valued 0-form" [i.e. $\binom{n}{0}$ tensor] $\boldsymbol{S}$, $\boldsymbol{d}\boldsymbol{S} \equiv \boldsymbol{\nabla}\boldsymbol{S}$.

Before proceeding further with the exterior (soon to be marked as "antisymmetric") differentiation of tensors, write down a formula (see exercise 14.5) for the exterior (antisymmetric) derivative of a product of forms:

$$\boldsymbol{d}(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) = (\boldsymbol{d}\boldsymbol{\alpha}) \wedge \boldsymbol{\beta} + (-1)^p \boldsymbol{\alpha} \wedge \boldsymbol{d}\boldsymbol{\beta}, \tag{14.13a}$$

where $\boldsymbol{\alpha}$ is a p-form and $\boldsymbol{\beta}$ is a q-form.

Now extend the exterior derivative from elementary forms to the exterior product of a tensor-valued p-form $\boldsymbol{S}$ with any ordinary q-form, $\boldsymbol{\beta}$; thus,

(3) defined in general

$$\boldsymbol{d}(\boldsymbol{S} \wedge \boldsymbol{\beta}) = \boldsymbol{d}\boldsymbol{S} \wedge \boldsymbol{\beta} + (-1)^p \boldsymbol{S} \wedge \boldsymbol{d}\boldsymbol{\beta}. \tag{14.13b}$$

This equation can be regarded as a general definition of the extended exterior derivative. For example, if $\boldsymbol{S}$ is a tensor-valued 2-form, $\boldsymbol{S} = S^{\alpha\beta}{}_{|\gamma\delta|}\boldsymbol{e}_\alpha \boldsymbol{e}_\beta\, \boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\delta$, then equation (14.13b) says

$$\boldsymbol{d}\boldsymbol{S} = \boldsymbol{d}[(\boldsymbol{e}_\alpha \boldsymbol{e}_\beta S^{\alpha\beta}{}_{|\gamma\delta|})(\boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\delta)] = \boldsymbol{d}(\boldsymbol{e}_\alpha \boldsymbol{e}_\beta S^{\alpha\beta}{}_{|\gamma\delta|}) \wedge (\boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\delta).$$

As another example, use (14.13b) to calculate $\boldsymbol{d}(\boldsymbol{u\sigma})$, where $\boldsymbol{u}$ is a vector-valued 0-form (vector) and $\boldsymbol{\sigma}$ is a scalar-valued 1-form (1-form):

$$\boldsymbol{d}(\boldsymbol{u\sigma}) = (\boldsymbol{du}) \wedge \boldsymbol{\sigma} + \boldsymbol{u}\,\boldsymbol{d\sigma}.$$

If one were following the practice of earlier chapters, one would have written $\boldsymbol{u} \otimes \boldsymbol{\sigma}$ where $\boldsymbol{u\sigma}$ appears here, $\boldsymbol{u} \otimes \boldsymbol{d\sigma}$ instead of $\boldsymbol{u}\,\boldsymbol{d\sigma}$, and $\boldsymbol{e}_\alpha \otimes \boldsymbol{e}_\beta$ instead of $\boldsymbol{e}_\alpha\boldsymbol{e}_\beta$. However, to avoid overcomplication in the notation, all such tensor product symbols are omitted here and hereafter.

Equations (14.12) and (14.13) do more than define the (extended) exterior derivative $\boldsymbol{d}$ and provide a way to use it in computations. They also allow one to define and calculate the antisymmetrized second derivatives, e.g., $\boldsymbol{d}^2\boldsymbol{v}$. The relation

$$\boldsymbol{d}^2\boldsymbol{v} = \mathcal{R}\boldsymbol{v}$$

where $\boldsymbol{v}$ is a vector will then introduce the "operator-valued" or "$\binom{1}{1}$-tensor valued" curvature 2-form $\mathcal{R}$. The notation of the extended exterior derivative puts a new look on the old apparatus of base vectors and parallel transport, and opens a way to calculate the curvature 2-form $\mathcal{R}$.

Let the vector field $\boldsymbol{v}$ be expanded in terms of some field of basis vectors $\boldsymbol{e}_\mu$; thus

$$\boldsymbol{v} = \boldsymbol{e}_\mu v^\mu.$$

Then the exterior derivative of this vector is

$$\boldsymbol{dv} = \boldsymbol{de}_\mu v^\mu + \boldsymbol{e}_\mu\,\boldsymbol{d}v^\mu.$$

Expand the typical vector-valued 1-form $\boldsymbol{de}_\mu$ in the form

Definition of $\boldsymbol{\omega}^\mu{}_\nu$

$$\boldsymbol{de}_\mu = \boldsymbol{e}_\nu \boldsymbol{\omega}^\nu{}_\mu. \tag{14.14}$$

Here the "components" $\boldsymbol{\omega}^\nu{}_\mu$ in the expansion of $\boldsymbol{de}_\mu$ are 1-forms. Recall from equation (10.13) that the typical $\boldsymbol{\omega}^\nu{}_\mu$ is related to the connection coefficients by

$$\boldsymbol{\omega}^\nu{}_\mu = \Gamma^\nu{}_{\mu\lambda}\boldsymbol{\omega}^\lambda. \tag{14.15}$$

Therefore the expansion of the "vector" (really, "vector-valued 1-form") is

$$\boldsymbol{dv} = \boldsymbol{e}_\mu(\boldsymbol{d}v^\mu + \boldsymbol{\omega}^\mu{}_\nu v^\nu). \tag{14.16}$$

Now differentiate once again to find

$$\begin{aligned}\boldsymbol{d}^2\boldsymbol{v} &= \boldsymbol{de}_\alpha \wedge (\boldsymbol{d}v^\alpha + \boldsymbol{\omega}^\alpha{}_\nu v^\nu)\\ &\quad + \boldsymbol{e}_\mu(\boldsymbol{d}^2v^\mu + \boldsymbol{d\omega}^\mu{}_\nu v^\nu - \boldsymbol{\omega}^\mu{}_\nu \wedge \boldsymbol{d}v^\nu)\\ &= \boldsymbol{e}_\mu(\boldsymbol{\omega}^\mu{}_\alpha \wedge \boldsymbol{d}v^\alpha + \boldsymbol{\omega}^\mu{}_\alpha \wedge \boldsymbol{\omega}^\alpha{}_\nu v^\nu\\ &\quad + \boldsymbol{d}^2v^\mu + \boldsymbol{d\omega}^\mu{}_\nu v^\nu - \boldsymbol{\omega}^\mu{}_\alpha \wedge \boldsymbol{d}v^\alpha).\end{aligned}$$

The simplifications made here use (1) the equation (14.14), for a second time; and (2) the product rule (14.13a), which introduced the minus sign in the last term, ready to cancel the first term. Now consider the term $\boldsymbol{d}^2 v^\mu$. Recall that any given component, for example, v^3, is an ordinary scalar function of position (as contrasted to $\boldsymbol{v}$ or $\boldsymbol{e}_3$ or $\boldsymbol{e}_3 v^3$). Therefore the standard exterior derivative (Chapter 4) as applied to a scalar function is all that $\boldsymbol{d}$ can mean in $\boldsymbol{d}^2 v^\mu$. But for the standard exterior derivative applied twice, one has automatically $\boldsymbol{d}^2 v^\mu = 0$ (Box 4.1, B; Box 4.4). This circumstance reduces the expansion for $\boldsymbol{d}^2 \boldsymbol{v}$ to the form

Curvature 2-forms $\mathcal{R}^\mu{}_\nu$:

$$\boldsymbol{d}^2 \boldsymbol{v} = \boldsymbol{e}_\mu \mathcal{R}^\mu{}_\nu v^\nu, \tag{14.17}$$

(1) in terms of $\boldsymbol{d}^2\boldsymbol{v}$

where the $\mathcal{R}^\mu{}_\nu$ are abbreviations for *the curvature 2-forms*

$$\mathcal{R}^\mu{}_\nu \equiv \boldsymbol{d}\boldsymbol{\omega}^\mu{}_\nu + \boldsymbol{\omega}^\mu{}_\alpha \wedge \boldsymbol{\omega}^\alpha{}_\nu. \tag{14.18}$$

(2) in terms of $\boldsymbol{\omega}^\mu{}_\nu$

Ordinarily, equation (14.18) surpasses in efficiency every other known method for calculating the curvature 2-forms.

The remarkable form of equation (14.17) deserves comment. On the left appear two $\boldsymbol{d}$'s, reminders that one has twice differentiated the vector field $\boldsymbol{v}$. But on the right, as the result of the differentiation, one has only the vector field $\boldsymbol{v}$ at the point in question, undifferentiated. How $\boldsymbol{v}$ varies from place to place enters not one whit in the answer. All that matters is how the geometry varies from place to place. Here is curvature coming into evidence. It comes into evidence free of any special features of the vector field $\boldsymbol{v}$, because the operation $\boldsymbol{d}^2$ is an antisymmetrized covariant derivative [compare equation (11.8) for this antisymmetrized covariant derivative in the previously developed abstract language, and see Boxes 11.2 and 11.6 for what is going on behind the scene expressed in the form of pictures]. In brief, the result of operating on $\boldsymbol{v}$ twice with $\boldsymbol{d}$ is an algebraic linear operation on $\boldsymbol{v}$; thus,

$$\boldsymbol{d}^2 \boldsymbol{v} = \mathcal{R}\boldsymbol{v}. \tag{14.19}$$

Tensor-valued curvature 2-form $\mathcal{R}$

Here $\mathcal{R}$ is an abbreviation for the "$\binom{1}{1}$-tensor valued 2-form,"

$$\mathcal{R} = \boldsymbol{e}_\mu \otimes \boldsymbol{\omega}^\nu \mathcal{R}^\mu{}_\nu. \tag{14.20}$$

If $\boldsymbol{d}$ is a derivative with a "slot in it" in which to insert the vector saying in what direction the differentiation is to proceed, then the $\boldsymbol{d}^2\boldsymbol{w}$ of $\boldsymbol{d}^2\boldsymbol{w} = \mathcal{R}\boldsymbol{w}$ has two slots and calls for two vectors, say, $\boldsymbol{u}$ and $\boldsymbol{v}$. These two vectors define the plane in which the antisymmetrized exterior derivative of (14.19) is to be evaluated (change in $\boldsymbol{w}$ upon going around the elementary route defined by $\boldsymbol{u}$ and $\boldsymbol{v}$ and coming back to its starting point; Boxes 11.6 and 11.7). To spell out explicitly this insertion of vectors into slots, return first to a simpler context, and see the exterior derivative of a 1-form (itself a 2-form) "evaluated" for a bivector $\boldsymbol{u} \wedge \boldsymbol{v}$ ("count of honeycomblike cells of the 2-form over the parallelogram-shaped domain defined by the two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$"), and see the result of the evaluation (exercise 14.6) expressed as a commutator,

$$\langle \boldsymbol{d\alpha}, \boldsymbol{u} \wedge \boldsymbol{v}\rangle = \partial_{\boldsymbol{u}}\langle \boldsymbol{\alpha}, \boldsymbol{v}\rangle - \partial_{\boldsymbol{v}}\langle \boldsymbol{\alpha}, \boldsymbol{u}\rangle - \langle \boldsymbol{\alpha}, [\boldsymbol{u}, \boldsymbol{v}]\rangle. \tag{14.21}$$

This result generalizes itself to a tensor-valued 1-form $\boldsymbol{S}$ of any rank in an obvious way; thus,

$$\langle \boldsymbol{dS}, \boldsymbol{u} \wedge \boldsymbol{v}\rangle = \boldsymbol{\nabla}_{\boldsymbol{u}}\langle \boldsymbol{S}, \boldsymbol{v}\rangle - \boldsymbol{\nabla}_{\boldsymbol{v}}\langle \boldsymbol{S}, \boldsymbol{u}\rangle - \langle \boldsymbol{S}, [\boldsymbol{u}, \boldsymbol{v}]\rangle. \tag{14.22}$$

Apply this result to the vector-valued 1-form $\boldsymbol{S} = \boldsymbol{dw}$. Recall the expression for a directional derivative, $\langle \boldsymbol{dw}, \boldsymbol{u}\rangle = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w}$. Thus find the result

$$\begin{aligned}\langle \boldsymbol{d}^2\boldsymbol{w}, \boldsymbol{u} \wedge \boldsymbol{v}\rangle &= \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{w} - \boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{w} - \boldsymbol{\nabla}_{[\boldsymbol{u},\boldsymbol{v}]}\boldsymbol{w}\\ &= \mathcal{R}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{w},\end{aligned} \tag{14.23}$$

Relation of curvature 2-form $\mathcal{R}$ to curvature operator $\mathcal{R}$

where $\mathcal{R}(\boldsymbol{u}, \boldsymbol{v})$ is the curvature operator defined already in Chapter 11 [equation (11.8)]. The conclusion is simple: the $\binom{1}{1}$-tensor-valued 2-form $\mathcal{R}$ of (14.19), evaluated on the bivector ("parallelogram") $\boldsymbol{u} \wedge \boldsymbol{v}$, is identical with the curvature operator $\mathcal{R}(\boldsymbol{u}, \boldsymbol{v})$ introduced previously; thus

$$\langle \mathcal{R}, \boldsymbol{u} \wedge \boldsymbol{v}\rangle = \mathcal{R}(\boldsymbol{u}, \boldsymbol{v}). \tag{14.24}$$

Now go from the language of abstract operators to a language that begins to make components show up. Substitute on the left the expression (14.20) and on the right the value of the curvature operator from (11.11); and rewrite (14.24) in the form

$$\boldsymbol{e}_\mu \otimes \boldsymbol{\omega}^\nu\langle \mathcal{R}^\mu{}_\nu, \boldsymbol{u} \wedge \boldsymbol{v}\rangle = \boldsymbol{e}_\mu \otimes \boldsymbol{\omega}^\nu R^\mu{}_{\nu\alpha\beta}u^\alpha v^\beta.$$

Compare and conclude that the typical individual curvature 2-form is given by the formula

Relation of $\mathcal{R}$ to components of ***Riemann***

$$\mathcal{R}^\mu{}_\nu = R^\mu{}_{\nu|\alpha\beta|}\boldsymbol{\omega}^\alpha \wedge \boldsymbol{\omega}^\beta \tag{14.25}$$

(sum over α, β, restricted to $\alpha < \beta$; so each index pair occurs only once).

Equation (14.25) provides the promised packaging of 21 curvature components into six curvature 2-forms; and equation (14.18) provides the quick means to calculate these curvature 2-forms. It is not necessary to take the key calculational equations (14.18) on faith, or to master the extended exterior derivative to prove or use them. Not one mention of any $\boldsymbol{d}$ do they make except the standard exterior $\boldsymbol{d}$ of Chapter 4. These key equations, moreover, can be verified in detail (exercise 14.8) by working in a coordinate frame. One adopts basis 1-forms $\boldsymbol{\omega}^\alpha = \boldsymbol{d}x^\alpha$. One goes on to use $\boldsymbol{\omega}^\mu{}_\nu = \Gamma^\mu{}_{\nu\lambda}\,\boldsymbol{d}x^\lambda$ from equation (14.15). In this way one obtains the "standard formula for the curvature" [equation (11.12) and equation (3) of Box 14.2] by standard methods.

In summary, the calculus of forms and exterior derivatives reduces the

$$\Gamma^\mu{}_{\alpha\beta} \longrightarrow R^\mu{}_{\nu\alpha\beta}$$

calculation to the

$$\boldsymbol{\omega}^\mu{}_\nu \longrightarrow \mathcal{R}^\mu{}_\nu$$

computation. Now look at the other link in the chain that leads from metric to curvature. It used to be

$$g_{\mu\nu} \longrightarrow \Gamma^\mu{}_{\alpha\beta}.$$

It now reduces to the calculation of "connection 1-forms"; thus

$$g_{\mu\nu} \longrightarrow \boldsymbol{\omega}^{\mu}{}_{\nu}.$$

Two principles master this first step in the curvature computation: (1) the symmetry of the covariant derivative; and (2) its compatibility with the metric. Condition (1), symmetry, appears in hidden guise in the principle

Symmetry of covariant derivative:

$$\boldsymbol{d}^2\mathscr{P} = 0. \tag{14.26}$$

(1) expressed as $\boldsymbol{d}^2\mathscr{P} = 0$

Here the notation "$\mathscr{P}$ for point" comes straight out of Cartan. He thought of a vector as defined by the movement of one point to another point infinitesimally close to it. To write $\boldsymbol{d}\mathscr{P}$ was therefore to take the "derivative of a point" [make a construction with a "point deleted" (tail of vector) and "point reinserted nearby" (tip of vector)]. The direction of the derivative $\boldsymbol{d}$ in $\boldsymbol{d}\mathscr{P}$ is indefinite. In other words, $\boldsymbol{d}\mathscr{P}$ contains a "slot." Only when one inserts into this slot a definite vector $\boldsymbol{v}$ does $\boldsymbol{d}\mathscr{P}$ give a definite answer for Cartan's vector. What is that vector that $\boldsymbol{d}\mathscr{P}$ then gives? It is $\boldsymbol{v}$ itself. "The movement that is $\boldsymbol{v}$ tells the point $\mathscr{P}$ to reproduce the movement that is $\boldsymbol{v}$"; or in concrete notation,

$$\langle \boldsymbol{d}\mathscr{P}, \boldsymbol{v}\rangle = \boldsymbol{v}. \tag{14.27}$$

Put the content of this equation into more formalistic terms. The quantity $\boldsymbol{d}\mathscr{P}$ is a $\binom{1}{1}$-tensor

$$\boldsymbol{d}\mathscr{P} = \boldsymbol{e}_{\mu}\boldsymbol{\omega}^{\mu}. \tag{14.28}$$

It is distinguished from the generic $\binom{1}{1}$-tensor

$$\boldsymbol{T} = \boldsymbol{e}_{\mu}T^{\mu}{}_{\nu}\boldsymbol{\omega}^{\nu}$$

by the special value of its components

$$T^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu}.$$

In this sense it deserves the name of "unit tensor." Insert this tensor in place of $\boldsymbol{S}$ into equation (14.22) and obtain the result

$$\langle \boldsymbol{d}^2\mathscr{P}, \boldsymbol{u} \wedge \boldsymbol{v}\rangle = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} - \boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{u} - [\boldsymbol{u}, \boldsymbol{v}] = 0. \tag{14.29}$$

The zero on the right is a restatement of equation (10.2a) or of "the closing of the vector diagram" in the picture called "symmetry of covariant differentiation" in Box 10.2. The vanishing of the righthand side for arbitrary $\boldsymbol{u}$ and $\boldsymbol{v}$ demands the vanishing of $\boldsymbol{d}^2\mathscr{P}$ on the left; and conversely, the vanishing of $\boldsymbol{d}^2\mathscr{P}$ demands the symmetry of the covariant derivative. The other principle basic to the forthcoming computations is "compatibility of covariant derivative with metric," as expressed in the form

$$\boldsymbol{d}(\boldsymbol{u}\cdot\boldsymbol{v}) = (\boldsymbol{du})\cdot\boldsymbol{v} + \boldsymbol{u}\cdot(\boldsymbol{dv}). \tag{14.30}$$

It is essential here to ascribe to the metric (the "dot") a vanishing covariant derivative; thus

$$\boldsymbol{d}(\cdot) = 0.$$

Capitalize on the symmetry and compatibility of the covariant derivative by using basis vectors (and where appropriate the basis 1-forms dual to these basis vectors) in equations (14.26) and (14.30). Thus from

$$\boldsymbol{d}\mathscr{P} = \boldsymbol{e}_\mu \boldsymbol{\omega}^\mu$$

compute

$$\begin{aligned} 0 = \boldsymbol{d}^2\mathscr{P} &= \boldsymbol{de}_\mu \wedge \boldsymbol{\omega}^\mu + \boldsymbol{e}_\mu \boldsymbol{d\omega}^\mu \\ &= \boldsymbol{e}_\mu(\boldsymbol{\omega}^\mu{}_\nu \wedge \boldsymbol{\omega}^\nu + \boldsymbol{d\omega}^\mu), \end{aligned}$$

and conclude that the coefficient of $\boldsymbol{e}_\mu$ must vanish; or

(2) expressed as $\boldsymbol{d\omega}^\mu + \boldsymbol{\omega}^\mu{}_\nu \wedge \boldsymbol{\omega}^\nu = 0$

$$0 = \boldsymbol{d\omega}^\mu + \boldsymbol{\omega}^\mu{}_\nu \wedge \boldsymbol{\omega}^\nu \qquad \text{("symmetry")}. \tag{14.31a}$$

Next, into (14.30) in place of the general $\boldsymbol{u}$ and $\boldsymbol{v}$ insert the specific $\boldsymbol{e}_\mu$ and $\boldsymbol{e}_\nu$, respectively, and find

Compatibility of $\boldsymbol{g}$ and $\boldsymbol{\nabla}$ expressed as $\boldsymbol{d}g_{\mu\nu} = \boldsymbol{\omega}_{\mu\nu} + \boldsymbol{\omega}_{\nu\mu}$

$$\boldsymbol{d}g_{\mu\nu} = \boldsymbol{\omega}_{\mu\nu} + \boldsymbol{\omega}_{\nu\mu} \qquad \text{("compatibility")}, \tag{14.31b}$$

where

$$\boldsymbol{\omega}_{\mu\nu} \equiv g_{\mu\alpha}\boldsymbol{\omega}^\alpha{}_\nu = \Gamma_{\mu\nu\alpha}\boldsymbol{\omega}^\alpha. \tag{14.31c}$$

In equations (14.31) one has the connection between metric and connection forms expressed in the most compact way.

§14.6. COMPUTATION OF CURVATURE USING EXTERIOR DIFFERENTIAL FORMS

Method of curvature 2-forms in 4 steps:

(1) select metric and frame

(2) calculate connection 1-forms $\boldsymbol{\omega}^\mu{}_\nu$

The use of differential forms for the computation of curvature is illustrated in Box 14.5. This section outlines the method. There are three main steps: compute $\boldsymbol{\omega}^\mu{}_\nu$; compute $\mathscr{R}^\mu{}_\nu$; and compute $G^\mu{}_\nu$. More particularly, first select a metric and a frame. Thereby fix the basis forms $\boldsymbol{\omega}^\mu = L^\mu{}_{\alpha'}\,\boldsymbol{d}x^{\alpha'}$ and the metric components $g_{\mu\nu}$ in $\boldsymbol{ds}^2 = g_{\mu\nu}\boldsymbol{\omega}^\mu \otimes \boldsymbol{\omega}^\nu$. Then determine the connection forms $\boldsymbol{\omega}^\mu{}_\nu$, and determine them uniquely, as solutions of the equations

$$0 = \boldsymbol{d\omega}^\mu + \boldsymbol{\omega}^\mu{}_\nu \wedge \boldsymbol{\omega}^\nu, \tag{14.31a}$$

$$\boldsymbol{d}g_{\mu\nu} = \boldsymbol{\omega}_{\mu\nu} + \boldsymbol{\omega}_{\nu\mu}. \tag{14.31b}$$

The "guess and check" method of finding a solution to these equations (described and illustrated in Box 14.5) is often quick and easy. [Exercise (14.7) shows that a solution always exists by showing that the Christoffel formula (14.36) is the unique solution in coordinate frames.] It is usually most convenient to use an orthonormal frame with $g_{\mu\nu} = \eta_{\mu\nu}$ (or some other simple frame where $g_{\mu\nu} = \text{const}$, e.g., a null frame). Then $\boldsymbol{d}g_{\mu\nu} = 0$ and equation (14.31b) shows that $\boldsymbol{\omega}_{\mu\nu} = -\boldsymbol{\omega}_{\nu\mu}$. Therefore there are only six $\boldsymbol{\omega}_{\mu\nu}$ for which to solve in four dimensions.

(*continued on page 358*)

Box 14.5 CURVATURE COMPUTED USING EXTERIOR DIFFERENTIAL FORMS (METRIC FOR FRIEDMANN COSMOLOGY)

The Friedmann metric

$$ds^2 = -dt^2 + a^2(t)[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)]$$

(Box 27.1) represents a spacetime where each constant-t hypersurface is a three-dimensional hypersphere of proper circumference $2\pi a(t)$. An orthonormal basis is easily found in this spacetime; thus,

$$\boldsymbol{ds}^2 = -(\boldsymbol{\omega}^{\hat{t}})^2 + (\boldsymbol{\omega}^{\hat{\chi}})^2 + (\boldsymbol{\omega}^{\hat{\theta}})^2 + (\boldsymbol{\omega}^{\hat{\phi}})^2,$$

where

$$\begin{aligned}\boldsymbol{\omega}^{\hat{t}} &= \boldsymbol{d}t,\\ \boldsymbol{\omega}^{\hat{\chi}} &= a\,\boldsymbol{d}\chi,\\ \boldsymbol{\omega}^{\hat{\theta}} &= a\sin\chi\,\boldsymbol{d}\theta,\\ \boldsymbol{\omega}^{\hat{\phi}} &= a\sin\chi\sin\theta\,\boldsymbol{d}\phi.\end{aligned} \tag{1}$$

A. Connection Computation

Equation (14.31b) gives, since $\boldsymbol{d}g_{\mu\nu} = \boldsymbol{d}\eta_{\mu\nu} = 0$, just

$$\boldsymbol{\omega}_{\mu\nu} = -\boldsymbol{\omega}_{\nu\mu}; \tag{2}$$

so there are only six 1-forms $\boldsymbol{\omega}_{\mu\nu}$ to be found. Turn to the second basic equation (14.31a). The game now is to *guess* a solution (because this is so often quicker than using systematic methods) to the equations $0 = \boldsymbol{d}\boldsymbol{\omega}^\mu + \boldsymbol{\omega}^\mu{}_\nu \wedge \boldsymbol{\omega}^\nu$ in which the $\boldsymbol{\omega}^\nu$ and thus also $\boldsymbol{d}\boldsymbol{\omega}^\mu$ are known, and $\boldsymbol{\omega}^\mu{}_\nu$ are unknown. The solution $\boldsymbol{\omega}^\mu{}_\nu$ is known to be unique; so guessing (if it leads to any answer) can only give the right answer.

Proceed from the simplest such equation. From $\boldsymbol{\omega}^{\hat{t}} = \boldsymbol{d}t$, compute

$$\boldsymbol{d}\boldsymbol{\omega}^{\hat{t}} = 0.$$

Compare this with $\boldsymbol{d}\boldsymbol{\omega}^{\hat{t}} = -\boldsymbol{\omega}^{\hat{t}}{}_\mu \wedge \boldsymbol{\omega}^\mu$ or (since $\boldsymbol{\omega}^{\hat{t}}{}_{\hat{t}} = -\boldsymbol{\omega}_{\hat{t}\hat{t}} = 0$, by $\boldsymbol{\omega}_{\mu\nu} = -\boldsymbol{\omega}_{\nu\mu}$)

$$\boldsymbol{d}\boldsymbol{\omega}^{\hat{t}} = -\boldsymbol{\omega}^{\hat{t}}{}_k \wedge \boldsymbol{\omega}^k = 0.$$

This equation could be satisfied by having $\boldsymbol{\omega}^{\hat{t}}{}_k \propto \boldsymbol{\omega}^k$, or in more complicated ways with cancelations among different terms, or more simply by $\boldsymbol{\omega}^{\hat{t}}{}_k = 0$. Proceed, not

Box 14.5 (*continued*)

looking for trouble, until some non-zero $\boldsymbol{\omega}^{\mu}{}_{\nu}$ is required. From $\boldsymbol{\omega}^{\hat{\chi}} = a\,\boldsymbol{d}\chi$, find

$$\begin{aligned} \boldsymbol{d\omega}^{\hat{\chi}} &= \dot{a}\,\boldsymbol{d}t \wedge \boldsymbol{d}\chi \\ &= (\dot{a}/a)\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\chi}} = -(\dot{a}/a)\boldsymbol{\omega}^{\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{t}}. \end{aligned}$$

Compare this with

$$\begin{aligned} \boldsymbol{d\omega}^{\hat{\chi}} &= -\boldsymbol{\omega}^{\hat{\chi}}{}_{\mu} \wedge \boldsymbol{\omega}^{\mu} \\ &= -\boldsymbol{\omega}^{\hat{\chi}}{}_{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{t}} - \boldsymbol{\omega}^{\hat{\chi}}{}_{\hat{\theta}} \wedge \boldsymbol{\omega}^{\hat{\theta}} - \boldsymbol{\omega}^{\hat{\chi}}{}_{\hat{\phi}} \wedge \boldsymbol{\omega}^{\hat{\phi}}. \end{aligned}$$

Guess that $\boldsymbol{\omega}^{\hat{\chi}}{}_{\hat{t}} = (\dot{a}/a)\boldsymbol{\omega}^{\hat{\chi}}$ from the first term; and hope the other terms vanish. (Note that this allows $\boldsymbol{\omega}^{\hat{t}}{}_{\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{\chi}} = -\boldsymbol{\omega}_{\hat{t}\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{\chi}} = \boldsymbol{\omega}_{\hat{\chi}\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\chi}} = 0$ in the $\boldsymbol{d\omega}^{\hat{t}}$ equation.) Look at $\boldsymbol{\omega}^{\hat{\theta}} = a \sin\chi\, \boldsymbol{d\theta}$, and write

$$\begin{aligned} \boldsymbol{d\omega}^{\hat{\theta}} &= (\dot{a}/a)\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\theta}} + a^{-1}\cot\chi\, \boldsymbol{\omega}^{\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{\theta}} \\ &= -\boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{t}} - \boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{\chi}} - \boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{\phi}} \wedge \boldsymbol{\omega}^{\hat{\phi}}. \end{aligned}$$

Guess, consistent with previously written equations, that

$$\begin{aligned} \boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{t}} &= \boldsymbol{\omega}^{\hat{t}}{}_{\hat{\theta}} = (\dot{a}/a)\boldsymbol{\omega}^{\hat{\theta}}, \\ \boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{\chi}} &= -\boldsymbol{\omega}^{\hat{\chi}}{}_{\hat{\theta}} = a^{-1}\cot\chi\, \boldsymbol{\omega}^{\hat{\theta}}. \end{aligned}$$

Finally from

$$\begin{aligned} \boldsymbol{d\omega}^{\hat{\phi}} &= (\dot{a}/a)\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\phi}} + a^{-1}\cot\chi\, \boldsymbol{\omega}^{\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{\phi}} \\ &\quad + (a\sin\chi)^{-1}\cot\theta\, \boldsymbol{\omega}^{\hat{\theta}} \wedge \boldsymbol{\omega}^{\hat{\phi}} \\ &= -\boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{t}} - \boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{\chi}} - \boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{\theta}} \wedge \boldsymbol{\omega}^{\hat{\theta}}, \end{aligned}$$

deduce values of $\boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{t}}$, $\boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{\chi}}$, and $\boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{\theta}}$. These are not inconsistent with previous assumptions that terms like $\boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{\phi}} \wedge \boldsymbol{\omega}^{\hat{\phi}}$ vanish (in the $\boldsymbol{d\omega}^{\hat{\theta}}$ equation); so one has in fact solved $\boldsymbol{d\omega}^{\mu} = -\boldsymbol{\omega}^{\mu}{}_{\nu} \wedge \boldsymbol{\omega}^{\nu}$ for a set of connection forms $\boldsymbol{\omega}^{\mu}{}_{\nu}$, as follows:

$$\begin{aligned} \boldsymbol{\omega}^{k}{}_{\hat{t}} &= \boldsymbol{\omega}^{\hat{t}}{}_{k} = (\dot{a}/a)\boldsymbol{\omega}^{k}, \\ \boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{\chi}} &= -\boldsymbol{\omega}^{\hat{\chi}}{}_{\hat{\theta}} = a^{-1}\cot\chi\, \boldsymbol{\omega}^{\hat{\theta}} \\ &= \cos\chi\, \boldsymbol{d\theta}, \\ \boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{\chi}} &= -\boldsymbol{\omega}^{\hat{\chi}}{}_{\hat{\phi}} = a^{-1}\cot\chi\, \boldsymbol{\omega}^{\hat{\phi}} \\ &= \cos\chi \sin\theta\, \boldsymbol{d\phi}, \\ \boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{\theta}} &= -\boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{\phi}} = (a\sin\chi)^{-1}\cot\theta\, \boldsymbol{\omega}^{\hat{\phi}} \\ &= \cos\theta\, \boldsymbol{d\phi}. \end{aligned} \tag{3}$$

Of course, if these hit-or-miss methods of finding $\boldsymbol{\omega}^{\mu}{}_{\nu}$ do not work easily in some problem, one may simply use equations (14.32) and (14.33).

B. Curvature Computation

The curvature computation is a straightforward substitution of $\boldsymbol{\omega}^{\mu}{}_{\nu}$ from equations (3) above into equation (14.34), which is

$$\mathscr{R}^{\mu}{}_{\nu} = \boldsymbol{d\omega}^{\mu}{}_{\nu} + \boldsymbol{\omega}^{\mu}{}_{\alpha} \wedge \boldsymbol{\omega}^{\alpha}{}_{\nu}.$$

This equation is short enough that one can write out the sum

$$\mathscr{R}^{\hat{t}}{}_{\hat{\chi}} = \boldsymbol{d\omega}^{\hat{t}}{}_{\hat{\chi}} + \boldsymbol{\omega}^{\hat{t}}{}_{\hat{\theta}} \wedge \boldsymbol{\omega}^{\hat{\theta}}{}_{\hat{\chi}} + \boldsymbol{\omega}^{\hat{t}}{}_{\hat{\phi}} \wedge \boldsymbol{\omega}^{\hat{\phi}}{}_{\hat{\chi}}$$

in contrast to the ten terms in the corresponding $R = \partial\Gamma + \Gamma^2$ equation [equation (3) of Box 14.2]. *Warning!:* From $\boldsymbol{\omega}^{\hat{t}}{}_{\hat{\chi}} = (\dot{a}/a)\boldsymbol{\omega}^{\hat{\chi}}$, do not compute $\boldsymbol{d\omega}^{\hat{t}}{}_{\hat{\chi}} = (\dot{a}/a)^{\boldsymbol{\cdot}}\, \boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\chi}}$. Missing is the term $(\dot{a}/a)\, \boldsymbol{d\omega}^{\hat{\chi}}$. Instead write $\boldsymbol{\omega}^{\hat{t}}{}_{\hat{\chi}} = (\dot{a}/a)\boldsymbol{\omega}^{\hat{\chi}} = \dot{a}\, \boldsymbol{d}\chi$, and then find $\boldsymbol{d\omega}^{\hat{t}}{}_{\hat{\chi}} = \ddot{a}\, \boldsymbol{d}t \wedge \boldsymbol{d}\chi = (\ddot{a}/a)\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\chi}}$. With elementary care, then, in correctly substituting from (3) for the $\boldsymbol{\omega}^{\mu}{}_{\nu}$ in the formula for $\mathscr{R}^{\mu}{}_{\nu}$, one finds

$$\mathscr{R}^{\hat{t}}{}_{\hat{\chi}} = (\ddot{a}/a)\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\chi}},$$

and

$$\mathscr{R}^{\hat{\chi}}{}_{\hat{\theta}} = (1 + \dot{a}^2)a^{-2}\boldsymbol{\omega}^{\hat{\chi}} \wedge \boldsymbol{\omega}^{\hat{\theta}}.$$

This completes the computation of the $R^{\mu}{}_{\nu\alpha\beta}$, since in this isotopic model universe, all space directions in the orthonormal frame $\boldsymbol{\omega}^{\mu}$ are algebraically equivalent. One can therefore write

$$\begin{aligned} \mathscr{R}^{\hat{t}}{}_{k} &= (\ddot{a}/a)\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{k}, \\ \mathscr{R}^{k}{}_{\ell} &= a^{-2}(1 + \dot{a}^2)\boldsymbol{\omega}^{k} \wedge \boldsymbol{\omega}^{\ell}, \end{aligned} \tag{4}$$

for the complete list of $\mathscr{R}^{\mu}{}_{\nu}$. Specific components, such as

$$R^{\hat{t}}{}_{\hat{\chi}\hat{t}\hat{\chi}} = \ddot{a}/a, \qquad R^{\hat{t}}{}_{\hat{\chi}\hat{t}\hat{\theta}} = 0, \text{ etc.},$$

or

$$R^{\hat{t}}{}_{\hat{\theta}\hat{\chi}\hat{\phi}} = 0, \qquad R^{\hat{\theta}}{}_{\hat{\phi}\hat{\theta}\hat{\phi}} = a^{-2}(1 + \dot{a}^2),$$

are easily read out of this display of $\mathscr{R}^{\mu}{}_{\nu}$.

C. Contraction

From equations (14.7), find

$$G^{\hat{t}\hat{t}} = +3a^{-2}(1 + \dot{a}^2), \tag{5a}$$

$$G^{\hat{t}\hat{\chi}} = G^{\hat{t}\hat{\theta}} = G^{\hat{t}\hat{\phi}} = 0 = G^{\hat{\chi}\hat{\theta}} = G^{\hat{\theta}\hat{\phi}} = G^{\hat{\phi}\hat{\chi}}, \tag{5b}$$

$$G^{\hat{\chi}\hat{\chi}} = G^{\hat{\theta}\hat{\theta}} = G^{\hat{\phi}\hat{\phi}} = -[2a^{-1}\ddot{a} + a^{-2}(1 + \dot{a}^2)], \tag{5c}$$

and

$$R = -G^{\mu}{}_{\mu} = 6[a^{-1}\ddot{a} + a^{-2}(1 + \dot{a}^2)]. \tag{6}$$

2

If guessing is not easy, there is a systematic way to solve equations (14.31) in an orthonormal frame or in any other frame in which $\boldsymbol{d}g_{\mu\nu} = 0$. Compute the $\boldsymbol{d\omega}^\mu$ and arrange them in the format

$$\boldsymbol{d\omega}^\alpha = -c_{|\mu\nu|}{}^\alpha \boldsymbol{\omega}^\mu \wedge \boldsymbol{\omega}^\nu. \tag{14.32}$$

In this way display the 24 "commutation coefficients" $c_{\mu\nu}{}^\alpha$. These quantities enter into the formula

$$\boldsymbol{\omega}_{\mu\nu} = \frac{1}{2}(c_{\mu\nu\alpha} + c_{\mu\alpha\nu} - c_{\nu\alpha\mu})\boldsymbol{\omega}^\alpha \tag{14.33}$$

to provide the six $\boldsymbol{\omega}^\mu{}_\nu$ (exercise 14.12).

(3) calculate curvature 2-forms $\mathcal{R}^\mu{}_\nu$

Once the $\boldsymbol{\omega}_{\mu\nu}$ are known, one computes the curvature forms $\mathcal{R}^\mu{}_\nu$ (again only six in four dimensions, since $\mathcal{R}^{\mu\nu} = -\mathcal{R}^{\nu\mu}$) by use of the formula

$$\mathcal{R}^\mu{}_\nu = \boldsymbol{d\omega}^\mu{}_\nu + \boldsymbol{\omega}^\mu{}_\alpha \wedge \boldsymbol{\omega}^\alpha{}_\nu. \tag{14.34}$$

(4) calculate components of curvature tensors

Out of this tabulation, one reads the individual components of the curvature tensor by using the identification scheme

$$\mathcal{R}^{\mu\nu} = R^{\mu\nu}{}_{|\alpha\beta|}\boldsymbol{\omega}^\alpha \wedge \boldsymbol{\omega}^\beta. \tag{14.35}$$

The Einstein tensor $G^\mu{}_\nu$ is computed by scanning the $\mathcal{R}^{\mu\nu}$ display to find the appropriate $R^{\mu\nu}{}_{\alpha\beta}$ components for use in formulas (14.7).

EXERCISES

Exercise 14.5. EXTERIOR DERIVATIVE OF A PRODUCT OF FORMS

Establish equation (14.13a) by working up recursively from forms of lower order to forms of higher order. [*Hints:* Recall from equation (4.27) that for a p-form

$$\boldsymbol{\alpha} = \alpha_{|\mu_1\ldots\mu_p|}\boldsymbol{d}x^{\mu_1} \wedge \cdots \wedge \boldsymbol{d}x^{\mu_p},$$

the exterior derivative is defined by

$$\boldsymbol{d\alpha} = \frac{\partial \alpha_{|\mu_1\ldots\mu_p|}}{\partial x^{\mu_0}}\boldsymbol{d}x^{\mu_0} \wedge \boldsymbol{d}x^{\mu_1} \wedge \cdots \wedge \boldsymbol{d}x^{\mu_p}.$$

Applied to the product $\boldsymbol{\alpha} \wedge \boldsymbol{\beta}$ of two 1-forms, this formula gives

$$\begin{aligned}
\boldsymbol{d}(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) &= \boldsymbol{d}[(\alpha_\lambda\, \boldsymbol{d}x^\lambda) \wedge (\beta_\mu\, \boldsymbol{d}x^\mu)] \\
&= \boldsymbol{d}[(\alpha_\lambda \beta_\mu)(\boldsymbol{d}x^\lambda \wedge \boldsymbol{d}x^\mu)] \\
&= \frac{\partial(\alpha_\lambda\beta_\mu)}{\partial x^\kappa}\boldsymbol{d}x^\kappa \wedge \boldsymbol{d}x^\lambda \wedge \boldsymbol{d}x^\mu \\
&= \left(\frac{\partial\alpha_\lambda}{\partial x^\kappa}\boldsymbol{d}x^\kappa \wedge \boldsymbol{d}x^\lambda\right) \wedge \beta_\mu\,\boldsymbol{d}x^\mu - (a_\lambda\,\boldsymbol{d}x^\lambda) \wedge \left(\frac{\partial\beta_\mu}{\partial x^\kappa}\boldsymbol{d}x^\kappa \wedge \boldsymbol{d}x^\mu\right) \\
&= (\boldsymbol{d\alpha}) \wedge \boldsymbol{\beta} - \boldsymbol{\alpha} \wedge \boldsymbol{d\beta}.
\end{aligned}$$

Extend the reasoning to forms of higher order.]

Exercise 14.6. RELATIONSHIP BETWEEN EXTERIOR DERIVATIVE AND COMMUTATOR

Establish formula (14.21) by showing (a) that the righthand side is an *algebraic* linear function of $\boldsymbol{u}$ and an algebraic linear function of $\boldsymbol{v}$, and (b) that the equation holds when $\boldsymbol{u}$ and $\boldsymbol{v}$ are coordinate basis vectors $\boldsymbol{u} = \partial/\partial x^k$, $\boldsymbol{v} = \partial/\partial x^l$.

Exercise 14.7. CHRISTOFFEL FORMULA DERIVED FROM CONNECTION FORMS

In a coordinate frame $\boldsymbol{\omega}^\mu = \boldsymbol{d}x^\mu$, show that equation (14.31a) requires $\Gamma^\mu{}_{\alpha\beta} = \Gamma^\mu{}_{\beta\alpha}$, and that, with this symmetry established, equation (14.31b) gives an expression for $\partial g_{\mu\nu}/\partial x^\alpha$ which can be solved to give the Christoffel formula

$$\Gamma^\mu{}_{\alpha\beta} = \frac{1}{2} g^{\mu\nu}\left(\frac{\partial g_{\nu\alpha}}{\partial x^\beta} + \frac{\partial g_{\nu\beta}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right). \tag{14.36}$$

Exercise 14.8. RIEMANN-CHRISTOFFEL CURVATURE FORMULA RELATED TO CURVATURE FORMS

Substitute $\boldsymbol{\omega}^\mu{}_\nu = \Gamma^\mu{}_{\nu\lambda}\, \boldsymbol{d}x^\lambda$ into equation (14.18), and from the result read out, according to equation (14.25), the classical formula (3) of Box 14.2 for the components $R^\mu{}_{\nu\alpha\beta}$.

Exercise 14.9. MATRIX NOTATION FOR REVIEW OF CARTAN STRUCTURE EQUATIONS

Let $e \equiv (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ be a row matrix whose entries are the basis vectors, and let ω be a column of basis 1-forms $\boldsymbol{\omega}^\mu$. Similarly let $\Omega = \|\boldsymbol{\omega}^\mu{}_\nu\|$ and $\mathcal{R} = \|\mathcal{R}^\mu{}_\nu\|$ be square matrices with 1-form and 2-form entries. This gives a compact notation in which $\boldsymbol{d}\boldsymbol{e}_\mu = \boldsymbol{e}_\nu \boldsymbol{\omega}^\nu{}_\mu$ and $\boldsymbol{d}\mathcal{P} = \boldsymbol{e}_\mu \boldsymbol{\omega}^\mu$ read

$$\boldsymbol{d}e = e\Omega \text{ and } \boldsymbol{d}\mathcal{P} = e\omega, \tag{14.37}$$

respectively.

(a) From equations (14.37) and $\boldsymbol{d}^2\mathcal{P} = 0$, derive equation (14.31a) in the form

$$0 = \boldsymbol{d}\omega + \Omega \wedge \omega. \tag{14.38}$$

[Solution: $\boldsymbol{d}^2\mathcal{P} = \boldsymbol{d}e \wedge \omega + e\,\boldsymbol{d}\omega = e(\Omega \wedge \omega + \boldsymbol{d}\omega)$.]

(b) Compute $\boldsymbol{d}^2 e$ as motivation for definition (14.18), which reads

$$\mathcal{R} = \boldsymbol{d}\Omega + \Omega \wedge \Omega. \tag{14.39}$$

(c) From $\boldsymbol{d}^2\omega = 0$, deduce $\mathcal{R} \wedge \omega = 0$ and then decompress the notation to get the antisymmetry relation $R^\mu{}_{[\alpha\beta\gamma]} = 0$.

(d) Compute $\boldsymbol{d}\mathcal{R}$ from equation (14.39), and relate it to the Bianchi identity $R^\mu{}_{\nu[\alpha\beta;\gamma]} = 0$.

(e) Let $v = \{v^\mu\}$ be a column of functions; so $\boldsymbol{v} = ev = \boldsymbol{e}_\mu v^\mu$ is a vector field. Compute, in compact notation, $\boldsymbol{d}\boldsymbol{v}$ and $\boldsymbol{d}^2\boldsymbol{v}$ to show $\boldsymbol{d}^2\boldsymbol{v} = e\mathcal{R}v$ (which is equation 14.17).

Exercise 14.10. TRANSFORMATION RULES FOR CONNECTION FORMS IN COMPACT NOTATION

Using the notation of the previous exercise, write $e' = eA$ in place of $\boldsymbol{e}_{\mu'} = \boldsymbol{e}_\nu A^\nu{}_{\mu'}$, and similarly $\omega' = A^{-1}\omega$, to represent a change of frame. Show that $\boldsymbol{d}\mathcal{P} \equiv e\omega = e'\omega'$. Substitute $e' = eA$ in $\boldsymbol{d}e' = e'\Omega'$ to deduce the transformation law

$$\Omega' = A^{-1}\Omega A + A^{-1}\,\boldsymbol{d}A. \tag{14.40}$$

Rewrite this in decompressed notation for coordinate frames with $A^\nu{}_{\mu'} = \partial x^\nu/\partial x^{\mu'}$ as a formula of the form $\Gamma^{\mu'}{}_{\alpha'\beta'} = (?)$.

2

Exercise 14.11. SPACE IS FLAT IF THE CURVATURE VANISHES (see §11.5)

If coordinates exist in which all straight lines ($d^2x^\mu/d\lambda^2 = 0$) are geodesics, then one says the space is flat. Evidently all $\Gamma^\mu{}_{\alpha\beta}$ and $R^\mu{}_{\nu\alpha\beta}$ vanish in this case, by equation (14.8) and equation (3) in Box 14.2. Show conversely that, if $\mathcal{R} = 0$, then such coordinates exist. Use the results of the previous problem to find differential equations for a transformation A to a basis e' where $\Omega' = 0$. What are the conditions for complete integrability of these equations? [Note that $\boldsymbol{d}f_K = F_K(x, f)$ is completely integrable if $\boldsymbol{d}^2 f_K = 0$ modulo the original equations.] Why will the basis forms $\boldsymbol{\omega}^{\mu'}$ in this new frame be coordinate differentials $\boldsymbol{\omega}^{\mu'} = \boldsymbol{d}x^{\mu'}$?

Exercise 14.12. SYSTEMATIC COMPUTATION OF CONNECTION FORMS IN ORTHONORMAL FRAMES

Deduce equation (14.32) by applying equation (14.21) to basis vectors, using equations (8.14) to define $c_{\mu\nu}{}^\alpha$. Then show that, in an orthonormal frame (or any frames with $g_{\mu\nu} = \text{const}$), equation (14.33) provides a solution of equations (14.31), which define $\boldsymbol{\omega}^\mu{}_\nu$. [Compare also equation (8.24b).]

Exercise 14.13. SCHWARZSCHILD CURVATURE FORMS

Use the obvious orthonormal frame $\boldsymbol{\omega}^{\hat{t}} = e^\Phi\, \boldsymbol{d}t$, $\boldsymbol{\omega}^{\hat{r}} = e^\Lambda\, \boldsymbol{d}r$, $\boldsymbol{\omega}^{\hat{\theta}} = r\, \boldsymbol{d}\theta$, $\boldsymbol{\omega}^{\hat{\phi}} = r \sin\theta\, \boldsymbol{d}\phi$ for the Schwarzschild metric

$$ds^2 = -e^{2\Phi}\, dt^2 + e^{2\Lambda}\, dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{14.41}$$

in which Φ and Λ are functions of r only; and compute the curvature forms $\mathcal{R}^{\hat{\mu}}{}_{\hat{\nu}}$ and the Einstein tensor $G^{\hat{\mu}}{}_{\hat{\nu}}$ by the methods of Box 14.5. [*Answer:* $\mathcal{R}^{\hat{t}\hat{r}} = E\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{r}}$, $\mathcal{R}^{\hat{t}\hat{\theta}} = \bar{E}\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\theta}}$, $\mathcal{R}^{\hat{t}\hat{\phi}} = \bar{E}\boldsymbol{\omega}^{\hat{t}} \wedge \boldsymbol{\omega}^{\hat{\phi}}$, $\mathcal{R}^{\hat{\theta}\hat{\phi}} = F\boldsymbol{\omega}^{\hat{\theta}} \wedge \boldsymbol{\omega}^{\hat{\phi}}$, $\mathcal{R}^{\hat{\phi}\hat{r}} = \bar{F}\boldsymbol{\omega}^{\hat{\phi}} \wedge \boldsymbol{\omega}^{\hat{r}}$, $\mathcal{R}^{\hat{r}\hat{\theta}} = \bar{F}\boldsymbol{\omega}^{\hat{r}} \wedge \boldsymbol{\omega}^{\hat{\theta}}$, with

$$\begin{aligned}
E &= -e^{-2\Lambda}(\Phi'' + \Phi'^2 - \Phi'\Lambda'),\\
\bar{E} &= -\frac{1}{r}e^{-2\Lambda}\Phi',\\
F &= \frac{1}{r^2}(1 - e^{-2\Lambda}),\\
\bar{F} &= \frac{1}{r}e^{-2\Lambda}\Lambda';
\end{aligned} \tag{14.42}$$

and then

$$\begin{aligned}
G^{\hat{t}}_{\hat{t}} &= -(F + 2\bar{F}),\\
G^{\hat{r}}_{\hat{r}} &= -(F + 2\bar{E}),\\
G^{\hat{\theta}}_{\hat{\theta}} &= G^{\hat{\phi}}_{\hat{\phi}} = -(E + \bar{E} + \bar{F}),\\
G^{\hat{t}}_{\hat{r}} &= G^{\hat{t}}_{\hat{\theta}} = G^{\hat{t}}_{\hat{\phi}} = 0 = G^{\hat{r}}_{\hat{\theta}} = G^{\hat{r}}_{\hat{\phi}} = G^{\hat{\theta}}_{\hat{\phi}}.]
\end{aligned} \tag{14.43}$$

Exercise 14.14. MATRIX DISPLAY OF THE RIEMANN-TENSOR COMPONENTS

Use the symmetries of the Riemann tensor to justify displaying its components in an orthonormal frame in the form

$$R^{\hat{\mu}\hat{\nu}}{}_{\hat{\alpha}\hat{\beta}} = \begin{array}{c} 01\\ 02\\ 03\\ 23\\ 31\\ 12 \end{array} \left(\begin{array}{c:c} & \\ E & H \\ & \\ \hdashline & \\ -H^T & F \\ & \end{array}\right), \tag{14.44}$$

where the rows are labeled by index pairs $\hat{\mu}\hat{\nu} = 01, 02$, etc., as shown; and the columns $\hat{\alpha}\hat{\beta}$, similarly. Here E, F, and H are each 3×3 matrices with (why?)

$$E = E^T, \qquad F = F^T, \qquad \text{trace } H = 0, \tag{14.45}$$

where E^T means the transpose of E.

Exercise 14.15. RIEMANN MATRIX WITH VANISHING EINSTEIN TENSOR

Show that the empty-space Einstein equations $G^{\hat{\mu}}{}_{\hat{\nu}} = 0$ allow the matrix in equation (14.44) to be simplified to the form

$$R^{\hat{\mu}\hat{\nu}}{}_{\hat{\alpha}\hat{\beta}} = \left(\begin{array}{c|c} E & H \\ \hline -H & E \end{array}\right), \tag{14.46}$$

where now, in addition to the equality $E = F$ that this form implies, the further conditions

$$\text{trace } E = 0, \; H = H^T \tag{14.47}$$

hold.

Exercise 14.16. COMPUTATION OF CURVATURE FOR A PULSATING OR COLLAPSING STAR

Spherically symmetric motions of self-gravitating bodies are discussed in Chapters 26 and 32. A metric form often adopted in this situation is

$$ds^2 = -e^{2\Phi}\, dT^2 + e^{2\Lambda}\, dR^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{14.48}$$

where now Φ, Λ, and r are each functions of the two coordinates R and T. Compute the curvature 2-forms and the Einstein tensor for this metric, using the methods of Box 14.5. In the guessing of the $\boldsymbol{\omega}^{\mu}{}_{\nu}$, most of the terms will already be evident from the corresponding calculation in exercise 14.13. [*Answer,* in the obvious orthonormal frame $\boldsymbol{\omega}^{\hat{T}} = e^{\Phi}\, \boldsymbol{d}T$, $\boldsymbol{\omega}^{\hat{R}} = e^{\Lambda}\, \boldsymbol{d}R$, $\boldsymbol{\omega}^{\hat{\theta}} = r\, \boldsymbol{d}\theta$, $\boldsymbol{\omega}^{\hat{\phi}} = r \sin\theta\, \boldsymbol{d}\phi$:

$$\begin{aligned}
\mathcal{R}^{\hat{T}}{}_{\hat{R}} &= E\boldsymbol{\omega}^{\hat{T}} \wedge \boldsymbol{\omega}^{\hat{R}}, \\
\mathcal{R}^{\hat{T}}{}_{\hat{\theta}} &= \bar{E}\boldsymbol{\omega}^{\hat{T}} \wedge \boldsymbol{\omega}^{\hat{\theta}} + H\boldsymbol{\omega}^{\hat{R}} \wedge \boldsymbol{\omega}^{\hat{\theta}}, \\
\mathcal{R}^{\hat{T}}{}_{\hat{\phi}} &= \bar{E}\boldsymbol{\omega}^{\hat{T}} \wedge \boldsymbol{\omega}^{\hat{\phi}} + H\boldsymbol{\omega}^{\hat{R}} \wedge \boldsymbol{\omega}^{\hat{\phi}}, \\
\mathcal{R}^{\hat{\theta}}{}_{\hat{\phi}} &= F\boldsymbol{\omega}^{\hat{\theta}} \wedge \boldsymbol{\omega}^{\hat{\phi}}, \\
\mathcal{R}^{\hat{R}}{}_{\hat{\theta}} &= \bar{F}\boldsymbol{\omega}^{\hat{R}} \wedge \boldsymbol{\omega}^{\hat{\theta}} - H\boldsymbol{\omega}^{\hat{T}} \wedge \boldsymbol{\omega}^{\hat{\theta}}, \\
\mathcal{R}^{\hat{R}}{}_{\hat{\phi}} &= \bar{F}\boldsymbol{\omega}^{\hat{R}} \wedge \boldsymbol{\omega}^{\hat{\phi}} - H\boldsymbol{\omega}^{\hat{T}} \wedge \boldsymbol{\omega}^{\hat{\phi}},
\end{aligned} \tag{14.49}$$

which, in the matrix display of exercise 14.14, gives

$$R^{\mu\nu}{}_{\alpha\beta} = \begin{array}{c} \left(\begin{array}{ccc|ccc} E & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \bar{E} & \cdot & \cdot & \cdot & H \\ \cdot & \cdot & \bar{E} & \cdot & -H & \cdot \\ \hline \cdot & \cdot & \cdot & F & \cdot & \cdot \\ \cdot & \cdot & H & \cdot & \bar{F} & \cdot \\ \cdot & -H & \cdot & \cdot & \cdot & \bar{F} \end{array}\right) \end{array} \begin{array}{l} \hat{T}\hat{R} \\ \hat{T}\hat{\theta} \\ \hat{T}\hat{\phi} \\ \hat{\theta}\hat{\phi} \\ \hat{\phi}\hat{R} \\ \hat{R}\hat{\theta} \end{array} . \tag{14.50}$$

Here

$$\begin{aligned}
E &= e^{-2\Phi}(\ddot{\Lambda} + \dot{\Lambda}^2 - \dot{\Lambda}\dot{\Phi}) - e^{-2\Lambda}(\Phi'' + \Phi'^2 - \Phi'\Lambda'),\\
\bar{E} &= \frac{1}{r}e^{-2\Phi}(\ddot{r} - \dot{r}\dot{\Phi}) - \frac{1}{r}e^{-2\Lambda}r'\Phi',\\
H &= \frac{1}{r}e^{-\Phi-\Lambda}(\dot{r}' - \dot{r}\Phi' - r'\dot{\Lambda}),\\
F &= \frac{1}{r^2}(1 - r'^2e^{-2\Lambda} + \dot{r}^2e^{-2\Phi}),\\
\bar{F} &= \frac{1}{r}e^{-2\Phi}\dot{r}\dot{\Lambda} + \frac{1}{r}e^{-2\Lambda}(r'\Lambda' - r'').
\end{aligned} \tag{14.51}$$

The Einstein tensor is

$$\begin{aligned}
G^{\hat{T}\hat{T}} &= -G^{\hat{T}}{}_{\hat{T}} = F + 2\bar{F},\\
G^{\hat{T}\hat{R}} &= G^{\hat{T}}{}_{\hat{R}} = 2H,\\
G^{\hat{T}}{}_{\hat{\theta}} &= G^{\hat{T}}{}_{\hat{\phi}} = 0,\\
G^{\hat{R}}{}_{\hat{R}} &= -(2\bar{E} + F),\\
G^{\hat{\theta}}{}_{\hat{\theta}} &= G^{\hat{\phi}}{}_{\hat{\phi}} = -(E + \bar{E} + \bar{F}),\\
G^{\hat{R}}{}_{\hat{\theta}} &= G^{\hat{R}}{}_{\hat{\phi}} = G^{\hat{\theta}}{}_{\hat{\phi}} = 0.]
\end{aligned} \tag{14.52}$$

Exercise 14.17. BIANCHI IDENTITY IN $\boldsymbol{d}\mathcal{R} = 0$ FORM

Define the Riemann tensor as a bivector-valued 2-form,

$$\mathcal{R} = \frac{1}{2}\boldsymbol{e}_\mu \wedge \boldsymbol{e}_\nu \mathcal{R}^{\mu\nu}, \tag{14.53}$$

and evaluate $\boldsymbol{d}\mathcal{R}$ to make it manifest that $\boldsymbol{d}\mathcal{R} = 0$. Use

$$\mathcal{R}^{\mu\nu} = \boldsymbol{d\omega}^{\mu\nu} - \boldsymbol{\omega}^{\mu}{}_{\alpha} \wedge \boldsymbol{\omega}^{\nu\alpha}, \tag{14.54}$$

which is derived easily in an orthonormal frame (adequate for proving $\boldsymbol{d}\mathcal{R} = 0$), or (as a test of skill) in a general frame where $\mathcal{R}^{\mu\nu} = \mathcal{R}^{\mu}{}_{\alpha}g^{\alpha\nu}$ and (why?) $\boldsymbol{d}g^{\mu\nu} = -g^{\mu\alpha}(\boldsymbol{d}g_{\alpha\beta})g^{\beta\nu}$. [*Note:* only wedge products between forms (not those between vectors) count in fixing signs in the product rule (14.13) for $\boldsymbol{d}$.]

Exercise 14.18. LOCAL CONSERVATION OF ENERGY AND MOMENTUM: $\boldsymbol{d}{}^*\boldsymbol{T} = 0$ MEANS $\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$

Let the duality operator $*$, as defined for exterior differential forms in Box 4.1, act on the forms, *but not on the contravariant vectors,* which appear when the stress-energy tensor $\boldsymbol{T}$ or the Einstein tensor $\boldsymbol{G}$ is written as a mixed $\binom{1}{1}$ tensor:

$$\boldsymbol{T} = \boldsymbol{e}_\mu T^{\mu}{}_{\nu}\boldsymbol{\omega}^\nu$$

or

$$\boldsymbol{G} = \boldsymbol{e}_\mu G^{\mu}{}_{\nu}\boldsymbol{\omega}^\nu.$$

(a) Give an expression for ${}^*\boldsymbol{T}$ (or ${}^*\boldsymbol{G}$) expanded in terms of basis vectors and forms.

(b) Show that

$$^*\boldsymbol{T} = \boldsymbol{e}_\mu T^{\mu\nu}\, d^3\Sigma_\nu,$$

where $d^3\Sigma_\nu = \epsilon_{\nu|\alpha\beta\gamma|}\boldsymbol{\omega}^\alpha \wedge \boldsymbol{\omega}^\beta \wedge \boldsymbol{\omega}^\gamma$ [see Box 5.4 and equations (8.10)].

(c) Compute $\boldsymbol{d}\,^*\boldsymbol{T}$ using the generalized exterior derivative $\boldsymbol{d}$; find that

$$\boldsymbol{d}\,^*\boldsymbol{T} = \boldsymbol{e}_\mu T^{\mu\nu}{}_{;\nu}\sqrt{|g|}\,\boldsymbol{\omega}^0 \wedge \boldsymbol{\omega}^1 \wedge \boldsymbol{\omega}^2 \wedge \boldsymbol{\omega}^3.$$

2

CHAPTER 15

BIANCHI IDENTITIES AND THE BOUNDARY OF A BOUNDARY

§15.1. BIANCHI IDENTITIES IN BRIEF

This chapter is entirely Track 2.
As preparation, one needs to have covered (1) Chapter 4 (differential forms) and (2) Chapter 14 (computation of curvature).
In reading it, one will be helped by Chapters 9–11 and 13.
It is not needed as preparation for any later chapter, but it will be helpful in Chapter 17 (Einstein field equations).

Geometry gives instructions to matter, but how does matter manage to give instructions to geometry? Geometry conveys its instructions to matter by a simple handle: "pursue a world line of extremal lapse of proper time (geodesic)." What is the handle by which matter can act back on geometry? How can one identify the right handle when the metric geometry of Riemann and Einstein has scores of interesting features? Physics tells one what to look for: *a machinery of coupling between gravitation* (*spacetime curvature*) *and source* (matter; stress-energy tensor $\boldsymbol{T}$) *that will guarantee the automatic conservation of the source* ($\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$). Physics therefore asks mathematics: "What tensor-like feature of the geometry is automatically conserved?" Mathematics comes back with the answer: "The Einstein tensor." Physics queries, "How does this conservation come about?" Mathematics, in the person of Élie Cartan, replies, "Through the principle that 'the boundary of a boundary is zero'" (Box 15.1).

Actually, two features of the curvature are automatically conserved; or, otherwise stated, the curvature satisfies two Bianchi identities, the subject of this chapter. Both features of the curvature, both "geometric objects," lend themselves to representation in diagrams, moreover, diagrams that show in action the principle that "the boundary of a boundary is zero." In this respect, the geometry of spacetime shows a striking analogy to the field of Maxwell electrodynamics.

Identities and conservation of the source: electromagnetism and gravitation compared:

In electrodynamics there are four potentials that are united in the 1-form $\boldsymbol{A} \equiv A_\mu\, \boldsymbol{d}x^\mu$. Out of this quantity by differentiation follows the ***Faraday***, $\boldsymbol{F} = \boldsymbol{dA}$. This field satisfies the identity $\boldsymbol{dF} = 0$ (identity, yes; identity lending itself to the definition of a conserved source, no).

$\boldsymbol{dF} \equiv 0$

In gravitation there are ten potentials (metric coefficients $g_{\mu\nu}$) that are united in the metric tensor $\boldsymbol{g} = g_{\mu\nu}\, \boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu$. Out of this quantity by two differentiations follows the curvature operator

$$\mathcal{R} = \frac{1}{4} \boldsymbol{e}_\mu \wedge \boldsymbol{e}_\nu R^{\mu\nu}{}_{\alpha\beta}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta.$$

This curvature operator satisfies the Bianchi identity $\boldsymbol{d}\mathcal{R} = 0$, where now "$\boldsymbol{d}$" is a generalization of Cartan's exterior derivative, described more fully in Chapter 14 (again an identity, but again one that does not lend itself to the definition of a conserved source).

$\boldsymbol{d}\mathcal{R} \equiv 0$

In electromagnetism, one has to go to the dual, $^*\boldsymbol{F}$, to have any feature of the field that offers a handle to the source, $\boldsymbol{d}^*\boldsymbol{F} = 4\pi\, ^*\boldsymbol{J}$. The conservation of the source, $\boldsymbol{d}^*\boldsymbol{J} = 0$, appears as a consequence of the identity $\boldsymbol{dd}^*\boldsymbol{F} = 0$; or, by a rewording of the reasoning (Box 15.1), as a consequence of the vanishing of the boundary of a boundary.

$\boldsymbol{dd}^*\boldsymbol{F} \equiv 0$ plus Maxwell equations $\Longrightarrow \boldsymbol{d}^*\boldsymbol{J} = 0$

(*continued on page 370*)

Box 15.1 THE BOUNDARY OF A BOUNDARY IS ZERO

A. The Idea in Its 1-2-3-Dimensional Form

Begin with an oriented cube or approximation to a cube (3-dimensional).

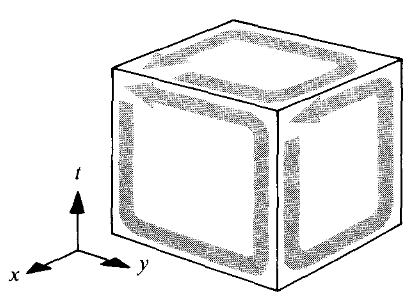
Its boundary is composed of six oriented faces, each two-dimensional. Orientation of each face is indicated by an arrow.

Boundary of any one oriented face consists of four oriented edges or arrows, each one-dimensional.

Every edge unites one face with another. No edge stands by itself in isolation.

"Sum" over all these edges, with due regard to sign. Find that any given edge is counted twice, once going one way, once going the other.

Conclude that the one-dimensional boundary of the two-dimensional boundary of the three-dimensional cube is identically zero.

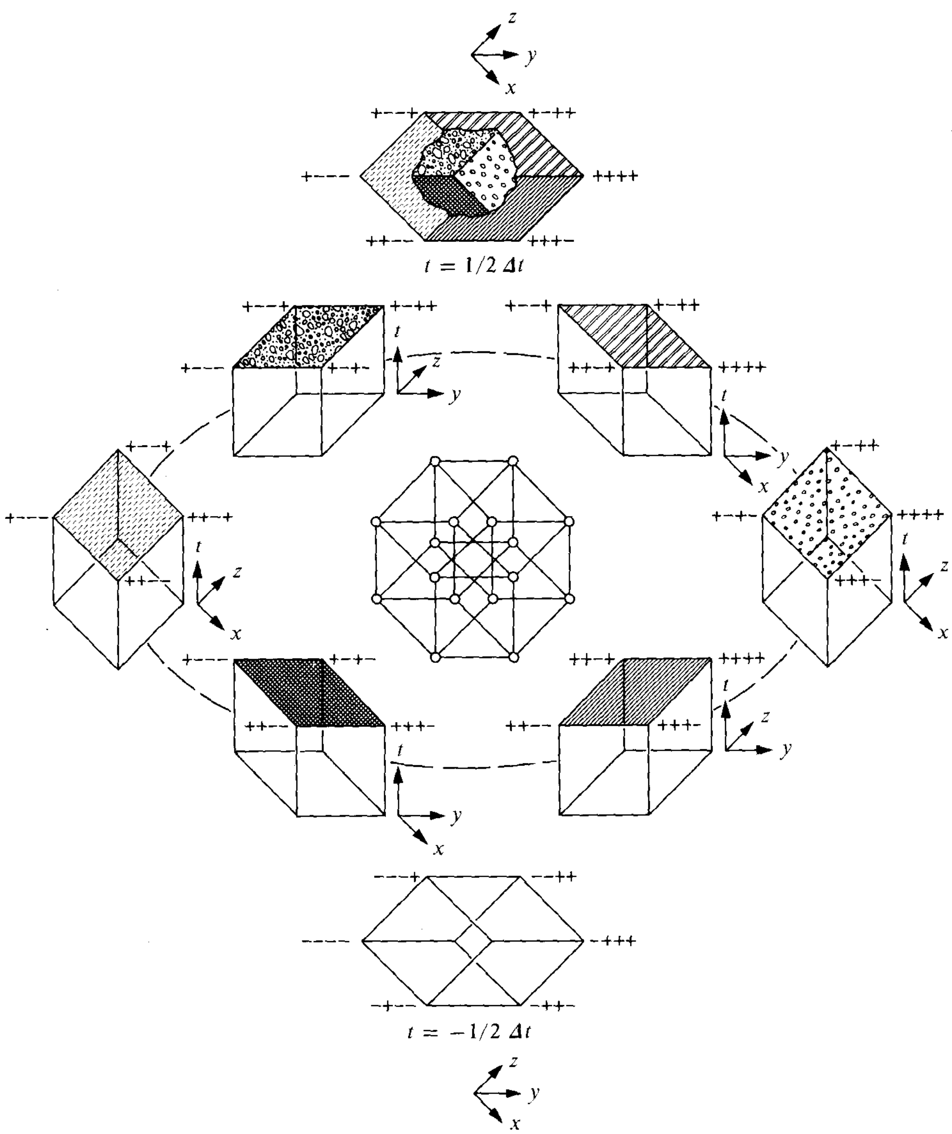


Box 15.1 (*continued*)

B. The Idea in Its 2-3-4-Dimensional Form

Begin with an oriented four-dimensional cube or approximation thereto. The coordinates of the typical corner of the four-cube may be taken to be $(t_0 \pm \frac{1}{2}\Delta t,\ x_0 \pm \frac{1}{2}\Delta x,\ y_0 \pm \frac{1}{2}\Delta y,\ z_0 \pm \frac{1}{2}\Delta z)$; and, accordingly, a sample corner itself, in an obvious abbreviation, is conveniently abbreviated $+--+$. There are 16 of these corners. Less complicated in appearance than the 4-cube itself are its three-dimensional faces, which are "exploded off of it" into the surrounding area of the diagram, where they can be inspected in detail.

The boundary of the 4-cube is composed of eight oriented hyperfaces, each of them three-dimensional (top hyperface with extension $\Delta x\,\Delta y\,\Delta z$, for example; a "front" hyperface with extension $\Delta t\,\Delta y\,\Delta z$; etc.)



Boundary of any one hyperface ("cube") consists of six oriented faces, each two-dimensional.

Every face (for example, the hatched face $\Delta x\ \Delta y$ in the lower lefthand corner) unites one hypersurface with another (the "3-cube side face" $\Delta t\ \Delta x\ \Delta y$ in the lower lefthand corner with the "3-cube top face" $\Delta x\ \Delta y\ \Delta z$, in this example). No face stands by itself in isolation. The three-dimensional boundary of the 4-cube exposes no 2-surface to the outside world. It is *faceless*.

"Sum" over all these faces, with due regard to orientation. Find any given face is counted twice, once with one orientation, once with the opposite orientation.

Conclude that the two-dimensional boundary of the three-dimensional boundary of the four-dimensional cube is identically zero.

C. The Idea in Its General Abstract Form

$\partial\partial = 0$ (the boundary of a boundary is zero).

D. Idea Behind Application to Gravitation and Electromagnetism

The one central point is a law of conservation (conservation of charge; conservation of momentum-energy).

The other central point is "automatic fulfillment" of this conservation law.

"Automatic conservation" requires that source not be an agent free to vary arbitrarily from place to place and instant to instant.

Source needs a tie to something that, while having degrees of freedom of its own, will cut down the otherwise arbitrary degrees of freedom of the source sufficiently to guarantee that the source automatically fulfills the conservation law. Give the name "field" to this something.

Define this field and "wire it up" to the source in such a way that the conservation of the source shall be an automatic consequence of the "zero boundary of a boundary." Or, more explicitly: Conservation demands no creation or destruction of source inside the four-dimensional cube shown in the diagram. Equivalently, integral of "creation events" (integral of $\boldsymbol{d}{}^{*}\boldsymbol{J}$ for electric charge; integral of $\boldsymbol{d}{}^{*}\boldsymbol{T}$ for energy-momentum) over this four-dimensional region is required to be zero.

Integral of creation over this four-dimensional region translates into integral of source density-current (${}^{*}\boldsymbol{J}$ or ${}^{*}\boldsymbol{T}$) over three-dimensional boundary of this region. This boundary consists of eight hyperfaces, each taken with due regard to orientation. Integral over upper hyperface ("$\Delta x\ \Delta y\ \Delta z$)" gives amount of source present at later moment; over lower hyperface gives amount of souce present at earlier moment; over such hyperfaces as "$\Delta t\ \Delta x\ \Delta y$" gives outflow of source over intervening period of time. Conservation demands that sum of these eight three-dimensional integrals shall be zero (details in Chapter 5).

Box 15.1 (*continued*)

Vanishing of this sum of three-dimensional integrals states the conservation requirement, but does not provide the machinery for "automatically" (or, in mathematical terms, "identically") meeting this requirement. For that, turn to principle that "boundary of a boundary is zero."

Demand that integral of source density-current over any oriented hyperface $\mathcal{V}$ (three-dimensional region; "cube") shall equal integral of field over faces of this "cube" (each face being taken with the appropriate orientation and the cube being infinitesimal):

$$4\pi \int_{\mathcal{V}} {}^*\boldsymbol{J} = \int_{\partial\mathcal{V}} {}^*\boldsymbol{F}; \qquad 8\pi \int_{\mathcal{V}} {}^*\boldsymbol{T} = \int_{\partial\mathcal{V}} \begin{pmatrix}\text{moment of}\\ \text{rotation}\end{pmatrix}.$$

Sum over the six faces of this cube and continue summing until the faces of all eight cubes are covered. Find that any given face (as, for example, the hatched face in the diagram) is counted twice, once with one orientation, once with the other ("boundary of a boundary is zero"). Thus is guaranteed the conservation of source: integral of source density-current over three-dimensional boundary of four-dimensional region is automatically zero, making integral of creation over interior of that four-dimensional region also identically zero.

Repeat calculation with boundary of that four-dimensional region slightly displaced in one locality [the "bubble differentiation" of Tomonaga (1946) and Schwinger (1948)], and conclude that conservation is guaranteed, not only in the four-dimensional region as a whole, but at every point within it, and, by extension, everywhere in spacetime.

E. Relation of Source to Field

One view: Source is primary. Field may have other duties, but its prime duty is to serve as "slave" of source. Conservation of source comes first; field has to adjust itself accordingly.

Alternative view: Field is primary. Field takes the responsibility of seeing to it that the source obeys the conservation law. Source would not know what to do in absence of the field, and would not even exist. Source is "built" from field. Conservation of source is consequence of this construction.

One model illustrating this view in an elementary context: Concept of "classical" electric charge as nothing but "electric lines of force trapped in the topology of a multiply connected space" [Weyl (1924b); Wheeler (1955); Misner and Wheeler (1957)].

On any view: Integral of source density-current over any three-dimensional region (a "cube" in simplified analysis above) equals integral of field over boundary of this region (the six faces of the cube above). No one has ever found any other way to understand the correlation between field law and conservation law.

F. Electromagnetism as a Model: How to "Wire Up" Source to Field to Give Automatic Conservation of Source Via "$\partial\partial = 0$" in Its 2-3-4-Dimensional Form

Conservation means zero creation of charge (zero creation in four-dimensional region Ω).

Conservation therefore demands zero value for integral of charge density-current over three-dimensional boundary of this volume; thus,

$$0 = \int_{\Omega} \frac{\partial J^{\mu}}{\partial x^{\mu}}\, d^4\Omega = \int_{\partial\Omega} J^{\mu}\, d^3\Sigma_{\mu}$$

in the Track-1 language of Chapters 3 and 5. Equivalently, in the coordinate-free abstract language of §§4.3-4.6, one has

$$0 = \int_{\Omega} \boldsymbol{d}{}^{*}\boldsymbol{J} = \int_{\partial\Omega} {}^{*}\boldsymbol{J},$$

where

$$\begin{aligned} {}^{*}\boldsymbol{J} &= {}^{*}J_{123}\, \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 + {}^{*}J_{023}\, \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 \\ &\quad + {}^{*}J_{031}\, \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^3 \wedge \boldsymbol{d}x^1 + {}^{*}J_{012}\, \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \end{aligned}$$

("eggcrate-like structure" of the 3-form of charge-density and current-density).

Fulfill this conservation requirement automatically ("identically") through the principle that "the boundary of a boundary is zero" by writing $4\pi\, {}^{*}\boldsymbol{J} = \boldsymbol{d}{}^{*}\boldsymbol{F}$; thus,

$$4\pi \int_{\partial\Omega} {}^{*}\boldsymbol{J} = \int_{\partial\Omega} \boldsymbol{d}{}^{*}\boldsymbol{F} = \int_{\partial\partial\Omega(\text{zero!})} {}^{*}\boldsymbol{F} \equiv 0$$

or, in Track-1 language, write $4\pi J^{\mu} = F^{\mu\nu}{}_{;\nu}$, and have

$$4\pi \int_{\partial\Omega} J^{\mu}\, d^3\Sigma_{\mu} = \int_{\partial\Omega} F^{\mu\nu}{}_{;\nu}\, d^3\Sigma_{\mu} = \int_{\partial\partial\Omega(\text{zero!})} F^{\mu\alpha}\, d^2\Sigma_{\mu\alpha} \equiv 0.$$

In other words, half of Maxwell's equations in their familiar flat-space form,

$$\operatorname{div} \boldsymbol{E} = \boldsymbol{\nabla}\cdot\boldsymbol{E} = 4\pi\rho, \qquad \operatorname{curl} \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{B} = \dot{\boldsymbol{E}} + 4\pi\boldsymbol{J},$$

"wire up" the source to the field in such a way that the law of conservation of source follows directly from "$\partial\partial\Omega = 0$."

G. Electromagnetism Also Employs "$\partial\partial = 0$" in its 1-2-3-Dimensional Form ("No Magnetic Charge")

Magnetic charge is linked with field via $4\pi\boldsymbol{J}_{\text{mag}} = \boldsymbol{d}\boldsymbol{F}$ (see point **F** above for translation of this compact Track-2 language into equivalent Track-1 terms). Absence of

Box 15.1 (*continued*)

any magnetic charge says that integral of $\boldsymbol{J}_{\text{mag}}$ over any 3-volume $\mathcal{V}$ is necessarily zero; or ("integration by parts," generalized Stokes theorem)

$$0 = \int_{\mathcal{V}} \boldsymbol{dF} = \int_{\partial\mathcal{V}} \boldsymbol{F} = \begin{pmatrix}\text{total magnetic flux} \\ \text{exiting through } \partial\mathcal{V}\end{pmatrix}.$$

In order to satisfy this requirement "automatically," via principle that "the boundary of a boundary is zero," write $\boldsymbol{F} = \boldsymbol{dA}$ ("expression of field in terms of 4-potential"), and have

$$\int_{\partial\mathcal{V}} \boldsymbol{F} = \int_{\partial\mathcal{V}} \boldsymbol{dA} = \int_{\partial\partial\mathcal{V}\,(\text{zero!})} \boldsymbol{A} \equiv 0.$$

H. Structure of Electrodynamics in Outline Form

Duality →

Differentiation ↓

$\boldsymbol{A}$ (potential)

↓

$\boldsymbol{F}$ (field; ***Faraday***) $= \boldsymbol{dA}$ ⟶ $^*\boldsymbol{F}$ (dual field; ***Maxwell***)

↓ ↓

$\boldsymbol{dF} = 0$ (identity based on $\partial\partial = 0$) $\quad$ $\boldsymbol{d}^*\boldsymbol{F} = 4\pi{}^*\boldsymbol{J}$

↓

$\boldsymbol{d}^*\boldsymbol{J} = 0$ (expressed as an identity based on $\partial\partial = 0$)

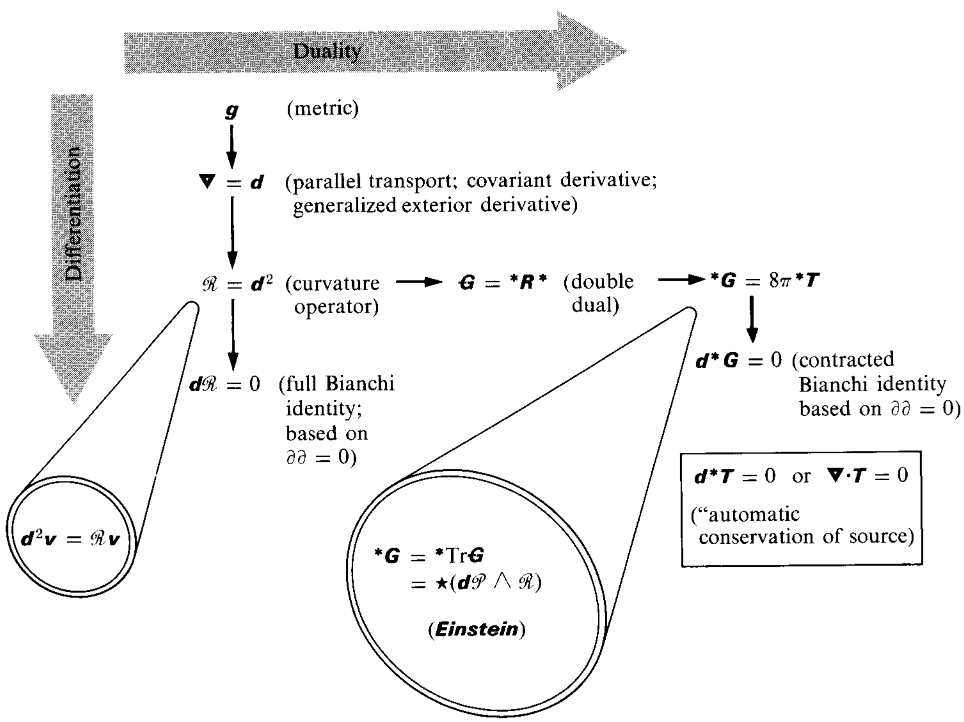
$\boldsymbol{d}^*\boldsymbol{J} = 0$ or $\boldsymbol{\nabla}\cdot\boldsymbol{J} = 0$ ("automatic" conservation of source)

In gravitation physics, one has to go to the "double dual" (two pairs of alternating indices, two places to take the dual) $\boldsymbol{\mathcal{G}} = {}^*\boldsymbol{R}^*$ of ***Riemann*** to have a feature of the field that offers a handle to the source:

$$\boldsymbol{G} = \text{Tr}\,\boldsymbol{\mathcal{G}} = \textbf{\textit{Einstein}} = 8\pi\boldsymbol{T} = 8\pi \times (\text{density of energy-momentum}).$$

The conservation of the source $\boldsymbol{T} \equiv \boldsymbol{e}_\mu T^\mu{}_\nu \boldsymbol{\omega}^\nu$ can be stated $\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$. But better suited for the present purpose is the form (see Chapter 14 and exercise 14.18)

I. Structure of Geometrodynamics in Outline Form



$$\boldsymbol{d}\,{}^*\boldsymbol{T} = 0,$$

$\boldsymbol{d}\,{}^*\boldsymbol{G} \equiv 0$ plus Einstein field equation $\Longrightarrow \boldsymbol{d}\,{}^*\boldsymbol{T} = 0$

where

$${}^*\boldsymbol{T} \equiv \boldsymbol{e}_\mu T^\mu{}_\nu({}^*\boldsymbol{\omega}^\nu) = \boldsymbol{e}_\mu T^{\mu\nu}\, d^3\Sigma_\nu.$$

This conservation law arises as a consequence of the "contracted Bianchi identity", $\boldsymbol{d}\,{}^*\boldsymbol{G} = 0$, again interpretable in terms of the vanishing of the boundary of a boundary.

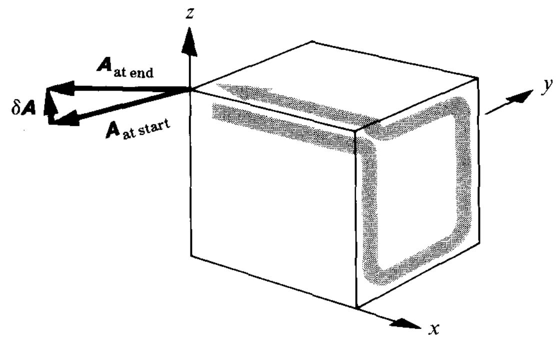


Figure 15.1.
Combine rotations associated with each of the six faces of the illustrated 3-volume and end up with zero net rotation ("full Bianchi identity"). Reason: Contribution of any face is measured by change in a test vector $\boldsymbol{A}$ carried in parallel transport around the perimeter of that face. Combine contributions of all faces and end up with each edge traversed twice, once in one direction, once in the other direction [boundary (here one-dimensional) of boundary (two-dimensional) of indicated three-dimensional figure is zero]. Detail: The vector $\boldsymbol{A}$, residing at the indicated site, is transported parallel to itself over to the indicated face, then carried around the perimeter of that face by parallel transport, experiencing in the process a rotation measured by the spacetime curvature associated with that face, then transported parallel to itself back to the original site. To the lowest relevant order of small quantities one can write

$$(\text{change in } \boldsymbol{A}) = -\Delta y\ \Delta z\ \mathcal{R}(\boldsymbol{e}_y, \boldsymbol{e}_z)\ \boldsymbol{A}$$

in operator notation; or in coordinate language,

$$-\delta A^\alpha = R^\alpha{}_{\beta yz}(\text{at } x + \Delta x)A^\beta\ \Delta y\ \Delta z.$$

§15.2. BIANCHI IDENTITY $d\mathcal{R} = 0$ AS A MANIFESTATION OF "BOUNDARY OF BOUNDARY = 0"

Bianchi identity, $\boldsymbol{d}\mathcal{R} \equiv 0$, interpreted in terms of parallel transport around the six faces of a cube.

Such is the story of the two Bianchi identities in outline form; it is now appropriate to fill in the details. Figure 15.1 illustrates the full Bianchi identity, $\boldsymbol{d}\mathcal{R} = 0$ (see exercise 14.17), saying in brief, "The sum of the curvature-induced rotations associated with the six faces of any elementary cube is zero." The change in a vector $\boldsymbol{A}$ associated with transport around the perimeter of the indicated face evaluated to the lowest relevant order of small quantities is given by

$$-\delta A^\alpha = R^\alpha{}_{\beta yz}\ (\text{at } x + \Delta x)A^\beta\ \Delta y\ \Delta z. \tag{15.1}$$

The opposite face gives a similar contribution, except that now the sign is reversed and the evaluation takes place at x rather than at $x + \Delta x$. The combination of the contributions from the two faces gives

$$\frac{\partial R^\alpha{}_{\beta yz}}{\partial x} A^\beta\ \Delta x\ \Delta y\ \Delta z, \tag{15.2}$$

when Riemann normal coordinates are in use. In such coordinates, the vanishing of the total $-\delta A^{\alpha}$ contributed by all six faces implies

$$R^{\alpha}{}_{\beta yz;x} + R^{\alpha}{}_{\beta zx;y} + R^{\alpha}{}_{\beta xy;z} = 0. \tag{15.3}$$

Here semicolons (covariant derivatives) can be and have been inserted instead of commas (ordinary derivatives), because the two are identical in the context of Riemann normal coordinates; and the covariant version (15.3) generalizes itself to arbitrary curvilinear coordinates. Turn from an *xyz* cube to a cube defined by any set of coordinate axes, and write Bianchi's identity in the form

$$R^{\alpha}{}_{\beta[\lambda\mu;\nu]} = 0. \tag{15.4}$$

(See exercise 14.17 for one reexpression of this identity in the abstract coordinate-independent form, $\boldsymbol{d}\mathcal{R} = 0$, and §15.3 for another.) This identity occupies much the same place in gravitation physics as that occupied by the identity $\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$ in electromagnetism:

$$F_{[\lambda\mu,\nu]} = F_{[\lambda\mu;\nu]} = 0. \tag{15.5}$$

§15.3. MOMENT OF ROTATION: KEY TO CONTRACTED BIANCHI IDENTITY

The contracted Bianchi identity, the identity that offers a "handle to couple to the source," was shown by Élie Cartan to deal with "moments of rotation" [Cartan (1928); Wheeler (1964b); Misner and Wheeler (1972)]. Moments are familiar in elementary mechanics. A rigid body will not remain at rest unless all the forces acting on it sum to zero:

$$\sum_i \boldsymbol{F}^{(i)} = 0. \tag{15.6}$$

Although necessary, this condition is not sufficient. The sum of the moments of these forces about some point $\mathcal{P}$ must also be zero:

$$\sum_i (\mathcal{P}^{(i)} - \mathcal{P}) \wedge \boldsymbol{F}^{(i)} = 0. \tag{15.7}$$

Exactly what point these moments are taken about happily does not matter, and this for a simple reason. The arbitrary point in the vector product (15.7) has for coefficient the quantity $\Sigma_i \boldsymbol{F}^{(i)}$, which already has been required to vanish. The situation is similar in the elementary cube of Figure 15.1. Here the rotation associated with a given face is the analog of the force $\boldsymbol{F}^{(i)}$ in mechanics. That the sum of these rotations vanishes when extended over all six faces of the cube is the analog of the vanishing of the sum of the forces $\boldsymbol{F}^{(i)}$.

What is the analog for curvature of the moment of the force that one encounters in mechanics? It is the *moment of the rotation associated with a given face of the*

2.

Net moment of rotation over all six faces of a cube:

(1) described

cube. The value of any individual moment depends on the reference point $\mathscr{P}$. However, the sum of these moments taken over all six faces of the cube will have a value independent of the reference point $\mathscr{P}$, for the same reason as in mechanics. Therefore $\mathscr{P}$ can be taken where one pleases, inside the elementary cube or outside it. Moreover, the cube may be viewed as a bit of a hypersurface sliced through spacetime. Therefore $\mathscr{P}$ can as well be off the slice as on it. It is only required that all distances involved be short enough that one obtains the required precision by calculating the moments and the sum of moments in a local Riemann-normal coordinate system. One thus arrives at a $\mathscr{P}$-independent totalized moment of rotation (not necessarily zero; gravitation is not mechanics!) associated with the cube in question.

(2) equated to integral of source, $\int {}^*\boldsymbol{T}$, over interior of cube

Now comes the magic of "the boundary of the boundary is zero." Identify this net moment of rotation of the cube, evaluated by summing individual moments of rotation associated with individual faces, with the integral of the source density-current (energy-momentum tensor ${}^*\boldsymbol{T}$) over the interior of the 3-cube. Make this identification not only for the one 3-cube, but for all eight 3-cubes (hyperfaces) that bound the four-dimensional cube in Box 15.1. Sum the integrated source density-current ${}^*\boldsymbol{T}$ not only for the one hyperface of the 4-cube, but for all eight hyperfaces. Thus have

$$\int_{\text{4-cube}} \begin{pmatrix} \text{source} \\ \text{creation} \\ \boldsymbol{d}{}^*\boldsymbol{T} \end{pmatrix} = \int_{\substack{\text{3-boundary} \\ \text{of this 4-cube}}} \begin{pmatrix} \text{source current-} \\ \text{density, } {}^*\boldsymbol{T} \end{pmatrix}$$

$$= \sum_{\substack{\text{these eight} \\ \text{bounding} \\ \text{3-cubes}}} \begin{pmatrix} \text{net moment of rotation} \\ \text{associated with speci-} \\ \text{fied cube} \end{pmatrix}$$

$$= \underbrace{\sum_{\substack{\text{eight} \\ \text{bounding} \\ \text{3-cubes}}} \sum_{\substack{\text{six faces} \\ \text{bounding} \\ \text{given 3-cube}}}}_{\text{(zero!)}} \begin{pmatrix} \text{moment of rotation} \\ \text{associated with specified} \\ \text{face of specified cube} \end{pmatrix}. \qquad (15.8)$$

(3) conserved

Let the moments of rotation, not only for the six faces of one cube, but for all the faces of all the cubes, be taken with respect to one and the same point $\mathscr{P}$. Recall (Box 15.1) that any given face joins two cubes or hyperfaces. It therefore appears twice in the count of faces, once with one orientation ("sense of circumnavigation in parallel transport to evaluate rotation") and once with the opposite orientation. Therefore the double sum vanishes identically (boundary of a boundary is zero!) This identity establishes existence of a new geometric object, a feature of the curvature, that is conserved, and therefore provides a handle to which to couple a source. The desired result has been achieved. Now to translate it into standard mathematics!

§15.4. CALCULATION OF THE MOMENT OF ROTATION

It remains to find the tensorial character and value of this conserved Cartan moment of rotation that appertains to any elementary 3-volume. The rotation associated with the front face $\Delta y\, \Delta z \boldsymbol{e}_y \wedge \boldsymbol{e}_z$ of the cube in Figure 15.1 will be represented by the bivector (4) evaluated

$$\begin{pmatrix}\text{rotation associated}\\ \text{with front } \Delta y\, \Delta z \text{ face}\end{pmatrix} = \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{yz}\, \Delta y\, \Delta z \tag{15.9}$$

located at $\mathscr{P}_{\text{front}} = (t - \frac{1}{2}\Delta t,\ x + \Delta x,\ y + \frac{1}{2}\Delta y,\ z + \frac{1}{2}\Delta z)$. This equation uses Riemann normal coordinates; indices enclosed by strokes, as in $|\lambda\mu|$, are summed with the restriction $\lambda < \mu$. The moment of this rotation with respect to the point $\mathscr{P}$ will be represented by the trivector

$$\begin{pmatrix}\text{moment of rotation}\\ \text{associated with}\\ \text{front } \Delta y\, \Delta z \text{ face}\end{pmatrix} = (\mathscr{P}_{\substack{\text{center}\\ \text{of front}\\ \text{face}}} - \mathscr{P}) \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{yz}\, \Delta y\, \Delta z. \tag{15.10}$$

Here neither $\mathscr{P}_{\text{center front}}$ nor $\mathscr{P}$ has any well-defined meaning whatsoever as a vector, but their difference is a vector in the limit of infinitesimal separation, $\Delta\mathscr{P} = \mathscr{P}_{\text{center front}} - \mathscr{P}$. With the back face a similar moment of rotation is associated, with the opposite sign, and with $\mathscr{P}_{\text{center front}}$ replaced by $\mathscr{P}_{\text{center back}}$. In the difference between the two terms, the factor $\mathscr{P}$ is of no interest, because one is already assured it will cancel out [Bianchi identity (15.4); analog of $\Sigma \boldsymbol{F}^{(i)} = 0$ in mechanics]. The difference $\mathscr{P}_{\text{center front}} - \mathscr{P}_{\text{center back}}$ has the value $\Delta x \boldsymbol{e}_x$. Summing over all six faces, one has

$$\begin{aligned}
&\begin{pmatrix}\text{net moment of}\\ \text{rotation associated}\\ \text{with cube or hyper-}\\ \text{face } \Delta x\, \Delta y\, \Delta z\end{pmatrix} = \\
&\qquad \boldsymbol{e}_x \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{yz}\, \Delta x\, \Delta y\, \Delta z \text{ (front and back)}\\
&\quad + \boldsymbol{e}_y \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{zx}\, \Delta y\, \Delta z\, \Delta x \text{ (sides)}\\
&\quad + \boldsymbol{e}_z \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{xy}\, \Delta z\, \Delta x\, \Delta y \text{ (top and bottom).}
\end{aligned} \tag{15.11}$$

This sum one recognizes as the value (on the volume element $\boldsymbol{e}_x \wedge \boldsymbol{e}_y \wedge \boldsymbol{e}_z\, \Delta x\, \Delta y\, \Delta z$) of the 3-form

$$\boldsymbol{e}_\nu \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta.$$

Moreover this 3-form is defined, and precisely defined, at a point, whereas (15.11), applying as it does to an extended region, does not lend itself to an analysis that is at the same time brief and precise. Therefore forego (15.11) in favor of the 3-form. Only remember, when it comes down to interpretation, that this 3-form is to be evaluated for the "cube" $\boldsymbol{e}_x \wedge \boldsymbol{e}_y \wedge \boldsymbol{e}_z\, \Delta x\, \Delta y\, \Delta z$. Now note that the "trivector-valued moment-of-rotation 3-form" can also be written as

(5) abstracted to give $\boldsymbol{d}\mathscr{P} \wedge \mathscr{R}$

$$\begin{pmatrix}\text{moment of}\\ \text{rotation}\end{pmatrix} = \boldsymbol{d}\mathscr{P} \wedge \mathscr{R} = \boldsymbol{e}_\nu \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta. \tag{15.12}$$

Here

$$\boldsymbol{d}\mathscr{P} = \boldsymbol{e}_\sigma\, \boldsymbol{d}x^\sigma \tag{15.13}$$

is Cartan's $\binom{1}{1}$ unit tensor. Also $\mathscr{R}$ is the curvature operator, treated as a bivector-valued 2-form:

$$\mathscr{R} = \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta. \tag{15.14}$$

Using the language of components as in (15.11), or the abstract language introduced in (15.12), one finds oneself dealing with a trivector. A trivector can be left a trivector, as, in quite another context, an element of 3-volume on a hypersurface in 4-space can be left as a trivector. However, there it is more convenient to take the dual representation, and speak of the element of volume as a vector. Denote by $\star$ a duality operation that acts only on contravariant vectors, trivectors, etc. (but *not* on forms). Then in a Lorentz frame one has $\star(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3) = \boldsymbol{e}_0$; but $\star(\boldsymbol{d}x^3) = \boldsymbol{d}x^3$. More generally,

$$\star(\boldsymbol{e}_\nu \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu) = \varepsilon_{\nu\lambda\mu}{}^\sigma \boldsymbol{e}_\sigma. \tag{15.15}$$

(6) abstracted to give $\star(\boldsymbol{d}\mathscr{P} \wedge \mathscr{R}) = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau$

In this notation, the "vector-valued moment-of-rotation 3-form" is

$$\begin{aligned}\begin{pmatrix}\text{moment}\\ \text{of rotation}\end{pmatrix} &= \star(\boldsymbol{d}\mathscr{P} \wedge \mathscr{R}) = \boldsymbol{e}_\sigma \varepsilon_{\nu\lambda\mu}{}^\sigma R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta\\ &= \boldsymbol{e}_\sigma ({}^*R)_\nu{}^\sigma{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta,\end{aligned}$$

or, in one more step,

$$\begin{pmatrix}\text{moment}\\ \text{of rotation}\end{pmatrix} = \star(\boldsymbol{d}\mathscr{P} \wedge \mathscr{R}) = \boldsymbol{e}_\sigma ({}^*R^*)_\nu{}^{\sigma\nu\tau}\, d^3\Sigma_\tau. \tag{15.16}$$

Here $d^3\Sigma_\tau$ is a notation for basis 3-forms, as in Box 5.4; thus,

$$\boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta = \varepsilon^{\nu\alpha\beta\tau}\, d^3\Sigma_\tau. \tag{15.17}$$

(In a local Lorentz frame, $\boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = d^3\Sigma_0$.)

Nothing is more central to the analysis of curvature than the formula (15.16). It starts with an element of 3-volume and ends up giving the moment of rotation in that 3-volume. The tensor that connects the starting volume with the final moment, the "contracted double-dual" of ***Riemann***, is so important that it deserves and receives a name of its own, $\boldsymbol{G} \equiv$ ***Einstein***; thus

$$(\text{Einstein})^{\sigma\tau} \equiv G^{\sigma\tau} = G_\nu{}^{\sigma\nu\tau} = ({}^*R^*)_\nu{}^{\sigma\nu\tau}. \tag{15.18}$$

This tensor received attention in §§13.5 and 14.2, and also in the examples at the end of Chapter 14. In terms of ***Einstein***, the connection between element of 3-volume and "vector-valued moment of rotation" is

$$\begin{pmatrix}\text{moment}\\ \text{of rotation}\end{pmatrix} = \star(\boldsymbol{d}\mathscr{P} \wedge \mathscr{R}) = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau. \tag{15.19}$$

The amount of "vector-valued moment of rotation" contained in the element of 3-volume $d^3\Sigma_\mu$ is identified by general relativity with the amount of energy-momentum contained in that 3-volume. However, defer this identification for now. Concentrate instead on the conservation properties of this moment of rotation. See them once in the formulation of integral calculus, as a consequence of the principle "$\partial\partial \equiv 0$." See them then a second time, in differential formulation, as a consequence of "$\boldsymbol{dd} \equiv 0$."

§15.5. CONSERVATION OF MOMENT OF ROTATION SEEN FROM "BOUNDARY OF A BOUNDARY IS ZERO"

Conservation of net moment of rotation:

The moment of rotation defines an automatically conserved quantity. In other words, the value of the moment of rotation for an elementary 3-volume $\Delta x\, \Delta y\, \Delta z$ after the lapse of a time Δt is equal to the value of the moment of rotation for the same 3-volume at the beginning of that time, corrected by the inflow of moment of rotation over the six faces of the 3-volume in that time interval (quantities proportional to $\Delta y\, \Delta z\, \Delta t$, etc.) Now verify this conservation of moment of rotation in the language of "the boundary of a boundary." Follow the pattern of equation (15.8), but translate the words into formulas, item by item. Evaluate the amount of moment of rotation created in the elementary 4-cube Ω, and find

(1) derived from "$\partial\partial = 0$"

$$\text{“creation”} \underset{\substack{\uparrow\\ \text{definition}}}{\equiv} \int \begin{pmatrix}\text{“creation of moment of}\\ \text{rotation” in the elementary}\\ \text{4-cube of spacetime } \Omega\end{pmatrix} \underset{\substack{\uparrow\\ \text{definition}}}{=} \int_\Omega \boldsymbol{d}{}^*\boldsymbol{G};$$

$$\int_\Omega \boldsymbol{d}{}^*\boldsymbol{G} \underset{\substack{\uparrow\\ \text{step 1}}}{=} \int_{\partial\Omega} {}^*\boldsymbol{G} \underset{\substack{\uparrow\\ \text{step 2}}}{=} \int_{\partial\Omega} \star(\boldsymbol{d}\mathscr{P} \wedge \mathscr{R}) \underset{\substack{\uparrow\\ \text{step 3}}}{=} \sum_{\substack{\text{the eight}\\ \text{3-cubes}\\ \text{that bound } \Omega}} \star \begin{pmatrix}\text{moment of rotation}\\ \displaystyle\int_{\text{3-cube}} (\boldsymbol{d}\mathscr{P} \wedge \mathscr{R})\\ \text{associated with}\\ \text{specified 3-cube}\end{pmatrix} \underset{\substack{\uparrow\\ \text{step 4}}}{=}$$

$$\underset{\substack{\uparrow\\ \text{step 4}}}{=} \sum_{\substack{\text{eight bounding}\\ \text{3-cubes}}} \sum_{\substack{\text{six faces bounding}\\ \text{specified 3-cube}}} \star \begin{pmatrix}\text{moment of rotation}\\ \displaystyle\int_{\text{face}} (\mathscr{P} \wedge \mathscr{R})\\ \text{associated with}\\ \text{specified face of}\\ \text{specified cube}\end{pmatrix} \underset{\substack{\uparrow\\ \text{step 5}}}{\equiv} 0. \tag{15.20}$$

Here step 1 is the theorem of Stokes. Step 2 is the identification established by (15.19) between the Einstein tensor and the moment of rotation. Step 3 breaks down the integral over the entire boundary $\partial\Omega$ into integrals over the individual 3-cubes that constitute this boundary. Moreover, in all these integrals, the star $\star$ is treated as a constant and taken outside the sign of integration. The reason for such treatment is simple: the duality operation $\star$ involves only the metric, and the metric is locally constant throughout the infinitesimal 4-cube over the boundary of which the integration extends. Step 4 uses the formula

$$\boldsymbol{d}(\mathcal{P} \wedge \mathcal{R}) = \boldsymbol{d}\mathcal{P} \wedge \mathcal{R} + \mathcal{P} \wedge \boldsymbol{d}\mathcal{R} = \boldsymbol{d}\mathcal{P} \wedge \mathcal{R} \tag{15.21}$$

and the theorem of Stokes to express each 3-cube integral as an integral of $\mathcal{P} \wedge \mathcal{R}$ over the two-dimensional boundary of that cube. The culminating step is 5. It has nothing to do with the integrand. It depends solely on the principle $\partial\partial \equiv 0$.

In brief, the conservation of moment of rotation follows from two circumstances. (1) The moment of rotation associated with any elementary 3-cube is by definition a net value, obtained by adding the six moments of rotation associated with the six faces of that cube. (2) When one sums these net values for all eight 3-cubes in (15.20), which are the boundary of the elementary 4-cube Ω, one counts the contribution of a given 2-face twice, once with one sign and once with the opposite sign. In virtue of the principle that "the boundary of a boundary is zero," the conservation of moment of rotation is thus an identity.

§15.6. CONSERVATION OF MOMENT OF ROTATION EXPRESSED IN DIFFERENTIAL FORM

(2) derived from "$\boldsymbol{dd} = 0$"

Every conservation law stated in integral form lends itself to restatement in differential form, and conservation of moment of rotation is no exception. The calculation is brief. Evaluate the generalized exterior derivative of the moment of rotation in three steps, and find that it vanishes; thus:

$$\begin{aligned} \boldsymbol{d}^*\boldsymbol{G} &= \boldsymbol{d}[\star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R})] \\ &= \star[\boldsymbol{d}(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R})] \\ &= \star[\boldsymbol{d}^2\mathcal{P} \wedge \mathcal{R} - \boldsymbol{d}\mathcal{P} \wedge \boldsymbol{d}\mathcal{R}] \\ &= 0 \end{aligned} \quad \begin{aligned} &\}\text{ step 1} \\ &\}\text{ step 2} \\ &\}\text{ step 3} \end{aligned}$$

Step 1 uses the relation $\boldsymbol{d}\star = \star\boldsymbol{d}$. The star duality and the generalized exterior derivative commute because when $\boldsymbol{d}$ is applied to a contravariant vector, it acts as a covariant derivative, and when $\star$ is applied to a covariant vector or 1-form, it is without effect. Step 2 applies the standard rule for the action of $\boldsymbol{d}$ on a product of tensor-valued forms [see equation (14.13b)]. Step 3 deals with two terms. The first term vanishes because the first factor in it vanishes; thus, $\boldsymbol{d}^2\mathcal{P} = 0$ [Cartan's equation of structure; expresses the "vanishing torsion" of the covariant derivative; see equation (14.26)]. The second term also vanishes, in this case, because the second factor in it vanishes; thus, $\boldsymbol{d}\mathcal{R} = 0$ (the full Bianchi identity). Thus briefly is conservation of moment of rotation established.

2

Box 15.2 THE SOURCE OF GRAVITATION AND THE MOMENT OF ROTATION: THE TWO KEY QUANTITIES AND THE MOST USEFUL MATHEMATICAL REPRESENTATIONS FOR THEM

	Energy-momentum as source of gravitation (curvature of space-time)	Moment of rotation as automatically conserved feature of the geometry
Representation as a vector-valued 3-form, a coordinate-independent geometric object	Machine to tell how much energy-momentum is contained in an elementary 3-volume: $^*\boldsymbol{T} = \boldsymbol{e}_\sigma T^{\sigma\tau}\, d^3\Sigma_\tau$ ("dual of stress-energy tensor")	Machine to tell how much net moment of rotation—expressed as a vector—is obtained by adding the six moments of rotation associated with the six faces of the elementary 3-cube: $\star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) = {}^*\boldsymbol{G} = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau$ ("dual of Einstein")
Representation as a $\binom{2}{0}$-tensor (also a coordinate independent geometric object)	Stress-energy tensor itself: $\boldsymbol{T} = \boldsymbol{e}_\sigma T^{\sigma\tau}\boldsymbol{e}_\tau$	Einstein itself: $\boldsymbol{G} = \boldsymbol{e}_\sigma G^{\sigma\tau}\boldsymbol{e}_\tau$
Representation in language of components (values depend on choice of coordinate system)	$T^{\sigma\tau}$	$G^{\sigma\tau}$
Conservation law in language of components	$T^{\sigma\tau}{}_{;\tau} = 0$	$G^{\sigma\tau}{}_{;\tau} \equiv 0$
Conservation in abstract language, for the $\binom{2}{0}$-tensor	$\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$	$\boldsymbol{\nabla}\cdot\boldsymbol{G} \equiv 0$
Conservation in abstract language, as translated into exterior derivative of the dual tensor (vector-valued 3-form)	$\boldsymbol{d}{}^*\boldsymbol{T} = 0$	$\boldsymbol{d}{}^*\boldsymbol{G} \equiv 0$ or $\boldsymbol{d}\star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) \equiv 0$
Same conservation law expressed in integral form for an element of 4-volume Ω	$\int_{\partial\Omega} {}^*\boldsymbol{T} = 0$	$\int_{\partial\Omega} {}^*\boldsymbol{G} \equiv 0$ or $\star\int_{\partial\Omega} (\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) \equiv 0$ or $\star\int_{\partial\partial\Omega} (\mathcal{P} \wedge \mathcal{R}) \equiv 0$

§15.7. FROM CONSERVATION OF MOMENT OF ROTATION TO EINSTEIN'S GEOMETRODYNAMICS: A PREVIEW

Einstein field equation "derived" from demand that (conservation of net moment of rotation) $\Rightarrow$ (conservation of source)

Mass, or mass-energy, is the source of gravitation. Mass-energy is one component of the energy-momentum 4-vector. Energy and momentum are conserved. The amount of energy-momentum in the element of 3-volume $d^3\boldsymbol{\Sigma}$ is

$$^*\boldsymbol{T} = \boldsymbol{e}_\sigma T^{\sigma\tau}\, d^3\Sigma_\tau \tag{15.22}$$

(see Box 15.2). Conservation of energy-momentum for an elementary 4-cube Ω expresses itself in the form

$$\int_{\partial\Omega} {}^*\boldsymbol{T} = 0. \tag{15.23}$$

This conservation is not an accident. According to Einstein and Cartan, it is "automatic"; and automatic, moreover, as a consequence of exact equality between energy-momentum and an automatically conserved feature of the geometry. What is this feature? It is the moment of rotation, which satisfies the law of automatic conservation,

$$\int_{\partial\Omega} {}^*\boldsymbol{G} = 0. \tag{15.24}$$

In other words, the conservation of momentum-energy is to be made geometric in character and automatic in action by the following prescription: *Identify the stress-energy tensor* (up to a factor 8π, or $8\pi G/c^4$, or other factor that depends on choice of units) *with the moment of rotation;* thus,

$$\star(\boldsymbol{d\mathcal{P}} \wedge \mathcal{R}) = {}^*\boldsymbol{G} = 8\pi\ {}^*\boldsymbol{T}; \tag{15.25}$$

or equivalently (still in the language of vector-valued 3-forms)

$$\begin{pmatrix}\text{moment of}\\ \text{rotation}\end{pmatrix} = \star(\boldsymbol{d\mathcal{P}} \wedge \mathcal{R}) = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau = 8\pi\boldsymbol{e}_\sigma T^{\sigma\tau}\, d^3\Sigma_\tau; \tag{15.26}$$

or, in the language of tensors,

$$\boldsymbol{G} = \boldsymbol{e}_\sigma G^{\sigma\tau}\boldsymbol{e}_\tau = 8\pi\boldsymbol{e}_\sigma T^{\sigma\tau}\boldsymbol{e}_\tau = 8\pi\boldsymbol{T}; \tag{15.27}$$

or, in the language of components,

$$G^{\sigma\tau} = 8\pi T^{\sigma\tau} \tag{15.28}$$

(Einstein's field equation; more detail, and more on the question of uniqueness, will be found in Chapter 17; see also Box 15.3). Thus simply is all of general relativity tied to the principle that the boundary of a boundary is zero. No one has ever discovered a more compelling foundation for the principle of conservation of momentum and energy. No one has ever seen more deeply into that action of matter on space, and space on matter, which one calls gravitation.

In summary, *the Einstein theory realizes the conservation of energy-momentum as the identity, "the boundary of a boundary is zero."*

EXERCISES

Exercise 15.1. THE BOUNDARY OF THE BOUNDARY OF A 4-SIMPLEX

In the analysis of the development in time of a geometry lacking all symmetry, when one is compelled to resort to a computer, one can, as one option, break up the 4-geometry into simplexes [four-dimensional analog of two-dimensional triangle, three-dimensional tetrahedron; vertices of "central simplex" conveniently considered to be at $(t, x, y, z) = (0, 1, 1, 1)$, $(0, 1, -1, -1)$, $(0, -1, 1, -1)$, $(0, -1, -1, 1)$, $(5^{1/2}, 0, 0, 0)$, for example], sufficiently numerous, and each sufficiently small, that the geometry inside each can be idealized as flat (Lorentzian), with all the curvature concentrated at the join between simplices (see discussion of dynamics of geometry via Regge calculus in Chapter 42). Determine ("give a mathematical

Box 15.3 OTHER IDENTITIES SATISFIED BY THE CURVATURE

(1) The source of gravitation is energy-momentum. (2) Energy-momentum is expressed by stress-energy tensor (or by its dual) as a vector-valued 3-form ("energy-momentum per unit 3-volume"). (3) This source is conserved (no creation in an elementary spacetime 4-cube).

These principles form the background for the probe in this chapter of the Bianchi identities. That is why two otherwise most interesting identities [Allendoerfer and Weil (1943); Chern (1955, 1962)] are dropped from attention. One deals with the 4-form

$$\Pi = \frac{1}{24\pi^2} g^{\alpha\gamma} g^{\beta\delta} \mathcal{R}_{\alpha\beta} \wedge \mathcal{R}_{\gamma\delta}, \tag{1}$$

and the other with the 4-form

$$\Gamma = \frac{1}{8\pi^2 |\det g_{\mu\nu}|^{1/2}} (\mathcal{R}_{12} \wedge \mathcal{R}_{30} + \mathcal{R}_{13} \wedge \mathcal{R}_{02} + \mathcal{R}_{10} \wedge \mathcal{R}_{23}). \tag{2}$$

Both quantities are built from the tensorial "curvature 2-forms"

$$\mathcal{R}_{\alpha\gamma} = \frac{1}{2} R_{\alpha\gamma\beta\delta}\, \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\delta. \tag{3}$$

The four-dimensional integral of either quantity over a four-dimensional region Ω has a value that (1) is a scalar, (2) is not identically equal to zero, (3) depends on the boundary of the region of spacetime over which the integral is extended, but (4) is independent of any changes made in the spacetime geometry interior to that surface (provided that these changes neither abandon the continuity nor change the connectivity of the 4-geometry in that region). Property (1) kills any possibility of identifying the integral, a scalar, with energy-momentum, a 4-vector. Property (2) kills it for the purpose of a conservation law, because it implies a non-zero creation in Ω.

Also omitted here is the Bel-Robinson tensor (see exercise 15.2), built bilinearly out of the curvature tensor, and other tensors for which see, e.g., Synge (1962).

One or all of these quantities may be found someday to have important physical content.

The integral of the 4-form Γ of equation (2) over the entire manifold gives a number, an integer, the so-called Euler-Poincaré characteristic of the manifold, whenever the integral and the integer are well-defined. This result is the four-dimensional generalization of the Gauss-Bonnet integral, widely known in the context of two-dimensional geometry:

$$\int \begin{pmatrix} \text{Riemannian scalar curvature} \\ \text{invariant (value } 2/a^2 \\ \text{for a sphere of radius } a) \end{pmatrix} g^{1/2}\, d^2x.$$

This integral has the value 8π for any closed, oriented, two-dimensional manifold with the topology of a 2-sphere, no matter how badly distorted; and the value 0 for any 2-torus, again no matter how rippled and twisted; and other equally specific values for other topologies.

description of") the boundary (three-dimensional) of such a simplex. Take one piece of this boundary and determine its boundary (two-dimensional). For one piece of this two-dimensional boundary, verify that there is at exactly one other place, and no more, in the book-keeping on the boundary of a boundary, another two-dimensional piece that cancels it ("facelessness" of the 3-boundary of the simplex).

2

Exercise 15.2. THE BEL-ROBINSON TENSOR [Bel (1958, 1959, 1962), Robinson (1959b), Sejnowski (1973); see also Pirani (1957) and Lichnerowicz (1962)].

Define the Bel-Robinson tensor by

$$T_{\alpha\beta\gamma\delta} = R_{\alpha\rho\gamma\sigma} R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma} + {}^{*}R_{\alpha\rho\gamma\sigma} {}^{*}R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma}. \tag{15.29}$$

Show that in empty spacetime this tensor can be rewritten as

$$T_{\alpha\beta\gamma\delta} = R_{\alpha\rho\gamma\sigma} R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma} + R_{\alpha\rho\delta\sigma} R_{\beta}{}^{\rho}{}_{\gamma}{}^{\sigma} - \frac{1}{8} g_{\alpha\beta} g_{\gamma\delta} R_{\rho\sigma\lambda\mu} R^{\rho\sigma\lambda\mu}. \tag{15.30a}$$

Show also that in empty spacetime

$$T^{\alpha}{}_{\beta\gamma\delta;\alpha} = 0, \tag{15.30b}$$

$T_{\alpha\beta\gamma\delta}$ is symmetric and traceless on all pairs of indices. (15.30c)

Discussion: It turns out that Einstein's "canonical energy-momentum pseudotensor" (§20.3) for the gravitational field in empty spacetime has a second derivative which, in a Riemann-normal coordinate system, is

$$t_{\mathrm{E}\alpha\beta,\gamma\delta} = -\frac{4}{9}\left(T_{\alpha\beta\gamma\delta} - \frac{1}{4} S_{\alpha\beta\gamma\delta}\right). \tag{15.31a}$$

Here $T_{\alpha\beta\gamma\delta}$ is the completely symmetric Bel-Robinson tensor, and $S_{\alpha\beta\gamma\delta}$ is defined by

$$S_{\alpha\beta\gamma\delta} \equiv R_{\alpha\delta\rho\sigma} R_{\beta\gamma}{}^{\rho\sigma} + R_{\alpha\gamma\rho\sigma} R_{\beta\delta}{}^{\rho\sigma} + \frac{1}{4} g_{\alpha\beta} g_{\gamma\delta} R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}. \tag{15.31b}$$

$S_{\alpha\beta\gamma\delta}$ appears in the empty-space covariant wave equation

$$\Delta R_{\alpha\beta\gamma\delta} \equiv -R_{\alpha\beta\gamma\delta;\mu}{}^{;\mu} + R_{\alpha\beta\rho\sigma} R_{\gamma\delta}{}^{\rho\sigma} + 2(R_{\alpha\rho\gamma\sigma} R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma} - R_{\alpha\rho\delta\sigma} R_{\beta}{}^{\rho}{}_{\gamma}{}^{\sigma}) = 0, \tag{15.31c}$$

where Δ is a variant of the Lichnerowicz-de Rham wave operator [Lichnerowicz (1964)], when one rewrites this wave equation as

$$\Box R_{\alpha\beta}{}^{\gamma\delta} \equiv R_{\alpha\beta}{}^{\gamma\delta}{}_{;\mu}{}^{;\mu} = 2S_{[\alpha}{}^{[\gamma}{}_{\beta]}{}^{\delta]}. \tag{15.31d}$$

PART **IV**

EINSTEIN'S GEOMETRIC THEORY OF GRAVITY

Wherein the reader is seduced into marriage with the most elegant temptress of all—Geometrodynamics—and learns from her the magic potions and incantations that control the universe.

CHAPTER 16

EQUIVALENCE PRINCIPLE AND MEASUREMENT OF THE "GRAVITATIONAL FIELD"

Rather than have one global frame with gravitational forces we have many local frames without gravitational forces.

STEPHEN SCHUTZ (1966)

§16.1. OVERVIEW

With the mathematics of curved spacetime now firmly in hand, one is tempted to rush headlong into a detailed study of Einstein's field equations. But such temptation must be resisted for a short time more. To grasp the field equations fully, one must first understand how the classical laws of physics change, or do not change, in the transition from flat spacetime to curved (§§16.2 and 16.3); and one must understand how the "gravitational field" (metric; covariant derivative; spacetime curvature; . . .) can be "measured" (§§16.4 and 16.5).

Purpose of this chapter

§16.2. THE LAWS OF PHYSICS IN CURVED SPACETIME

Wherever one is and whenever one probes, one finds that then and there one can introduce a local inertial frame in which all test particles move along straight lines. Moreover, this local inertial frame is also locally Lorentz: in it the velocity of light has its standard value, and light rays, like world lines of test particles, are straight. But physics is more, and the analysis of physics demands more than an account solely of the motions of test particles and light rays. What happens to Maxwell's equations, the laws of hydrodynamics, the principles of atomic structure, and all the rest of physics under the influence of "powerful gravitational fields"?

Einstein's equivalence principle

The answer is simple: *in any and every local Lorentz frame, anywhere and anytime in the universe, all the (nongravitational) laws of physics must take on their familiar special-relativistic forms.* Equivalently: there is no way, by experiments confined to infinitesimally small regions of spacetime, to distinguish one local Lorentz frame in one region of spacetime from any other local Lorentz frame in the same or any other region. This is Einstein's principle of equivalence in its strongest form—a principle that is compelling both philosophically and experimentally. (For the relevant experimental tests, see §38.6.)

Equivalence principle as tool to mesh nongravitational laws with gravity

The principle of equivalence has great power. With it one can generalize all the special relativistic laws of physics to curved spacetime. And the curvature need not be small. It may be as large as that in the center of a neutron star; as large as that at the edge of a black hole; arbitrarily large, in fact—or almost so. Only at the endpoint of gravitational collapse and in the initial instant of the "big bang," i.e., only at "singularities of spacetime," will there be a breakdown in the conditions needed for direct application of the equivalence principle (see §§28.3, 34.6, 43.3, 43.4, and chapter 44). Everywhere else the equivalence principle acts as a tool to mesh all the nongravitational laws of physics with gravity.

Example: Mesh the "law of local energy-momentum conservation," $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$, with gravity. *Solution:*

(1) The law in flat spacetime, written in abstract geometric form, reads

$$\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0. \tag{16.1a}$$

(2) Rewritten in a global Lorentz frame of flat spacetime, it reads

$$T^{\mu\nu}{}_{,\nu} = 0. \tag{16.1b}$$

(3) Application of equivalence principle gives same equation in local Lorentz frame of curved spacetime:

$$T^{\hat{\mu}\hat{\nu}}{}_{,\hat{\nu}} = 0 \text{ at origin of local Lorentz frame.} \tag{16.1c}$$

Because the connection coefficients vanish at the origin of the local Lorentz frame, this can be rewritten as

$$T^{\hat{\mu}\hat{\nu}}{}_{;\hat{\nu}} = 0 \text{ at origin of local Lorentz frame.} \tag{16.1d}$$

(4) The geometric law in curved spacetime, of which these are the local-Lorentz components, is

$$\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0; \tag{16.1e}$$

and its component formulation in any reference frame reads

$$T^{\mu\nu}{}_{;\nu} = 0. \tag{16.1f}$$

Compare the abstract geometric law (16.1e) in curved spacetime with the corresponding law (16.1a) in flat spacetime. They are identical! That this is not an accident one can readily see by tracing out the above four-step argument for any other law of physics (e.g., Maxwell's equation $\boldsymbol{\nabla} \cdot \boldsymbol{F} = 4\pi \boldsymbol{J}$). *The laws of physics, written in abstract geometric form, differ in no way whatsoever between curved spacetime and flat spacetime;* this is guaranteed by, and in fact is a mere rewording of, the equivalence principle.

Compare the component version of the law $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$, as written in an arbitrary frame in curved spacetime [equation (16.1f)], with the component version in a global Lorentz frame of flat spacetime [equation (16.1b)]. They differ in only one way: the comma (partial derivative; flat-spacetime gradient) is replaced by a semicolon (covariant derivative; curved-spacetime gradient). This procedure for rewriting the equations has universal application. *The laws of physics, written in component form, change on passage from flat spacetime to curved spacetime by a mere replacement of all commas by semicolons* (no change at all physically or geometrically; change due only to switch in reference frame from Lorentz to non-Lorentz!). This statement, like the nonchanging of abstract geometric laws, is nothing but a rephrased version of the equivalence principle.

"Comma-goes-to-semicolon" rule

The transition in formalism from flat spacetime to curved spacetime is a trivial process when performed as outlined above. But it is nontrivial in its implications. It meshes gravity with all the laws of physics. Gravity enters in an essential way through the covariant derivative of curved spacetime, as one sees clearly in the following exercise.

EXERCISES

Exercise 16.1. HYDRODYNAMICS IN A WEAK GRAVITATIONAL FIELD

(a) In §18.4 it will be shown that for a nearly Newtonian system, analyzed in an appropriate nearly global Lorentz coordinate system, the metric has the form

$$ds^2 = -(1 + 2\Phi)\,dt^2 + (1 - 2\Phi)(dx^2 + dy^2 + dz^2) \tag{16.2a}$$

where Φ is the Newtonian potential $(-1 \ll \Phi < 0)$. Consider a nearly Newtonian perfect fluid [stress-energy tensor

$$T^{\alpha\beta} = (\rho + p)u^\alpha u^\beta + pg^{\alpha\beta}, \qquad p \ll \rho; \tag{16.2b}$$

see Box 5.1 and §5.10] moving in such a spacetime with ordinary velocity

$$v^j \equiv dx^j/dt \ll 1. \tag{16.2c}$$

Show that the equations $T^{\mu\nu}{}_{;\nu} = 0$ for this system reduce to the familiar Newtonian law of mass conservation, and the Newtonian equation of motion for a fluid in a gravitational field:

$$\frac{d\rho}{dt} = -\rho\frac{\partial v^j}{\partial x^j}, \qquad \rho\frac{dv^j}{dt} = -\rho\frac{\partial\Phi}{\partial x^j} - \frac{\partial p}{\partial x^j}, \tag{16.3a}$$

where d/dt is the time derivative comoving with the matter

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + v^j\frac{\partial}{\partial x^j}. \tag{16.3b}$$

(b) Use these equations to calculate the pressure gradient in the Earth's atmosphere as a function of temperature and pressure. In the calculation, use the nonrelativistic relation $\rho = n_M \mu_M$, where n_M is the number density of molecules and μ_M is the mean rest mass per molecule; use the ideal-gas equation of state

$$p = n_M kT \qquad (k = \text{Boltzmann's constant});$$

and use the spherically symmetric form, $\Phi = -M/r$, for the Earth's Newtonian potential. If the pressure at sea level is 1.01×10^6 dynes/cm^2, what, approximately, is the pressure on top of Mount Everest (altitude 8,840 meters)? (Make a reasonable assumption about the temperature distribution of the atmosphere.)

Exercise 16.2. WORLD LINES OF PHOTONS

Show that in flat spacetime the conservation law for the 4-momentum of a freely moving photon can be written

$$\nabla_{\boldsymbol{p}} \boldsymbol{p} = 0. \tag{16.4a}$$

According to the equivalence principle, this equation must be true also in curved spacetime. Show that this means photons move along null geodesics of curved spacetime with affine parameter λ related to 4-momentum by

$$\boldsymbol{p} = d/d\lambda \tag{16.4b}$$

In exercise 18.6 this result will be used to calculate the deflection of light by the sun.

---

§16.3. FACTOR-ORDERING PROBLEMS IN THE EQUIVALENCE PRINCIPLE

Factor-ordering problems and coupling to curvature

On occasion in applying the equivalence principle to get from physics in flat spacetime to physics in curved spacetime one encounters "factor-ordering problems" analogous to those that beset the transition from classical mechanics to quantum mechanics.* *Example:* How is the equation (3.56) for the vector potential of electrodynamics to be translated into curved spacetime? If the flat-spacetime equation is written

$$-A^{\alpha,\mu}{}_{\mu} + A^{\mu}{}_{,\mu}{}^{\alpha} = 4\pi J^{\alpha},$$

then its transition ("comma goes to semicolon") reads

$$-A^{\alpha;\mu}{}_{\mu} + A^{\mu}{}_{;\mu}{}^{\alpha} = 4\pi J^{\alpha}. \tag{16.5}$$

However, if the flat-spacetime equation is written with two of its partial derivatives interchanged

$$-A^{\alpha,\mu}{}_{\mu} + A^{\mu,\alpha}{}_{\mu} = 4\pi J^{\alpha},$$

*For a discussion of quantum-mechanical factor-ordering problems, see, e.g., Merzbacher (1961), pp. 138–39 and 334–35; also Pauli (1934).

then its translation reads

$$-A^{\alpha;\mu}{}_{\mu} + A^{\mu;\alpha}{}_{\mu} = 4\pi J^{\alpha},$$

which can be rewritten

$$-A^{\alpha;\mu}{}_{\mu} + A^{\mu}{}_{;\mu}{}^{\alpha} + R^{\alpha}{}_{\mu}A^{\mu} = 4\pi J^{\alpha}. \tag{16.5$'$}$$

(Ricci tensor appears as result of interchanging covariant derivatives; see exercise 16.3.) Which equation is correct—(16.5) or (16.5′)? This question is nontrivial, just as the analogous factor-ordering problems of quantum theory are nontrivial. For rules-of-thumb that resolve this and most factor-ordering problems, see Box 16.1. These rules tell one that (16.5′) is correct and (16.5) is wrong (see Box 16.1 and §22.4).

EXERCISES

Exercise 16.3. NONCOMMUTATION OF COVARIANT DERIVATIVES

Let $\boldsymbol{B}$ be a vector field and $\boldsymbol{S}$ be a second-rank tensor field. Show that

$$B^{\mu}{}_{;\alpha\beta} = B^{\mu}{}_{;\beta\alpha} + R^{\mu}{}_{\nu\beta\alpha}B^{\nu} \tag{16.6a}$$

$$S^{\mu\nu}{}_{;\alpha\beta} = S^{\mu\nu}{}_{;\beta\alpha} + R^{\mu}{}_{\rho\beta\alpha}S^{\rho\nu} + R^{\nu}{}_{\rho\beta\alpha}S^{\mu\rho}. \tag{16.6b}$$

From equation (16.6a), show that

$$B^{\mu;\alpha}{}_{\mu} = B^{\mu}{}_{;\mu}{}^{\alpha} + R^{\alpha}{}_{\mu}B^{\mu}. \tag{16.6c}$$

[*Hint* for Track-1 calculation: Work in a local Lorentz frame, where $\Gamma^{\alpha}{}_{\beta\gamma} = 0$ but $\Gamma^{\alpha}{}_{\beta\gamma,\delta} \neq 0$; expand the lefthand side in terms of Christoffel symbols and partial derivatives; and use equation (8.44) for the Riemann tensor. An alternative Track-2 calculation notices that $\boldsymbol{\nabla}_{\beta}\boldsymbol{\nabla}_{\alpha}\boldsymbol{B}$ is *not* linear in $\boldsymbol{e}_{\alpha}$, and that $B^{\mu}{}_{;\alpha\beta}$ are *not* its components; but, rather, that

$$B^{\mu}{}_{;\alpha\beta} \equiv \underbrace{\boldsymbol{\nabla}\boldsymbol{\nabla}\boldsymbol{B}}_{\text{[Third-rank tensor]}}(\boldsymbol{\omega}^{\mu}, \boldsymbol{e}_{\alpha}, \boldsymbol{e}_{\beta}). \tag{16.7}$$

The calculation then proceeds as follows:

$$\begin{aligned}
\langle \boldsymbol{\omega}^{\mu}, \boldsymbol{\nabla}_{\beta}\boldsymbol{\nabla}_{\alpha}\boldsymbol{B}\rangle &= \langle \boldsymbol{\omega}^{\mu}, \boldsymbol{\nabla}_{\beta}(\boldsymbol{e}_{\alpha}\cdot\boldsymbol{\nabla}\boldsymbol{B})\rangle \\
&= \langle \boldsymbol{\omega}^{\mu}, (\boldsymbol{\nabla}_{\beta}\boldsymbol{e}_{\alpha})\cdot\boldsymbol{\nabla}\boldsymbol{B} + \boldsymbol{e}_{\alpha}\cdot(\boldsymbol{\nabla}_{\beta}\boldsymbol{\nabla}\boldsymbol{B})\rangle \\
&= \langle \boldsymbol{\omega}^{\mu}, \Gamma^{\nu}{}_{\alpha\beta}\boldsymbol{e}_{\nu}\cdot\boldsymbol{\nabla}\boldsymbol{B} + \boldsymbol{\nabla}\boldsymbol{\nabla}\boldsymbol{B}(\ldots, \boldsymbol{e}_{\alpha}, \boldsymbol{e}_{\beta})\rangle \\
&= B^{\mu}{}_{;\nu}\Gamma^{\nu}{}_{\alpha\beta} + B^{\mu}{}_{;\alpha\beta}.
\end{aligned}$$

Consequently

$$\begin{aligned}
B^{\mu}{}_{;\alpha\beta} - B^{\mu}{}_{;\beta\alpha} &= \langle \boldsymbol{\omega}^{\mu}, [\boldsymbol{\nabla}_{\beta}, \boldsymbol{\nabla}_{\alpha}]\boldsymbol{B}\rangle - B^{\mu}{}_{;\nu}(\Gamma^{\nu}{}_{\alpha\beta} - \Gamma^{\nu}{}_{\beta\alpha}) \\
&= \langle \boldsymbol{\omega}^{\mu}, [\boldsymbol{\nabla}_{\beta}, \boldsymbol{\nabla}_{\alpha}]\boldsymbol{B}\rangle - \langle \boldsymbol{\omega}^{\mu}, \boldsymbol{\nabla}_{(\boldsymbol{\nabla}_{\beta}\boldsymbol{e}_{\alpha} - \boldsymbol{\nabla}_{\alpha}\boldsymbol{e}_{\beta})}\boldsymbol{B}\rangle \\
&= \langle \boldsymbol{\omega}^{\mu}, ([\boldsymbol{\nabla}_{\beta}, \boldsymbol{\nabla}_{\alpha}] - \boldsymbol{\nabla}_{[\boldsymbol{e}_{\beta},\boldsymbol{e}_{\alpha}]})\boldsymbol{B}\rangle = \langle \boldsymbol{\omega}^{\mu}, \mathcal{R}(\boldsymbol{e}_{\beta}, \boldsymbol{e}_{\alpha})\boldsymbol{B}\rangle \\
&= R^{\mu}{}_{\nu\beta\alpha}B^{\nu},
\end{aligned}$$

in agreement with (16.6a). Note: because of slight ambiguity in the abstract notation, one must think carefully about each step in the above calculation. Component notation, by contrast, is completely unambiguous.]

(*continued on page 392*)

Box 16.1 FACTOR ORDERING AND CURVATURE COUPLING IN APPLICATIONS OF THE EQUIVALENCE PRINCIPLE

The Problem

In what order should derivatives be written when applying the "comma-goes-to-semicolon rule"? Interchanging derivatives makes no difference in flat spacetime, but in curved spacetime it produces terms that couple to curvature, e.g., $2B^{\alpha}{}_{;[\gamma\beta]} \equiv B^{\alpha}{}_{;\gamma\beta} - B^{\alpha}{}_{;\beta\gamma} = R^{\alpha}{}_{\mu\beta\gamma}B^{\mu}$ for any vector field (see exercise 16.3). Hence, the problem can be restated: *When must the comma-goes-to-semicolon rule be augmented by terms that couple to curvature?*

The Solution

There is no solution in general, but in most cases the following types of mathematical and physical reasoning resolve the problem unambiguously.

A. *Mathematically, curvature terms almost always arise from the noncommutation of covariant derivatives.* Consequently, one needs to worry about curvature terms in any equation that contains a double covariant derivative (e.g., $-A^{\alpha,\mu}{}_{\mu} + A^{\mu}{}_{,\mu}{}^{\alpha} = 4\pi J^{\alpha}$); or in any equation whose derivation from more fundamental laws involves double covariant derivatives (e.g. $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{S} = 0$ in Example B(3) below). But one can ignore curvature coupling everywhere else (e.g., in Maxwell's first-order equations).

B. *Coupling to curvature can surely not occur without some physical reason.* Therefore, if one applies the comma-goes-to-semicolon rule only to physically measurable quantities (e.g., to the electromagnetic field, but not to the vector potential), one can "intuit" whether coupling to curvature is likely. *Examples:*

(1) *Local energy-momentum conservation.* A coupling to curvature in the equations $T^{\alpha\beta}{}_{;\beta} = 0$—e.g., replacing them by $T^{\alpha\beta}{}_{;\beta} = R^{\alpha}{}_{\beta\gamma\delta}T^{\beta\gamma}u^{\delta}$—would not make sense at all. In a local inertial frame such terms as $R^{\alpha}{}_{\beta\gamma\delta}T^{\beta\gamma}u^{\delta}$ would be interpreted as forces produced at a single point by curvature. But it should not be possible to feel curvature except over finite regions (geodesic deviation, etc.)! Put differently, the second derivatives of the gravitational potential (metric) can hardly produce net forces at a point; they should only produce tidal forces!

(2) *Maxwell's equations* for the electromagnetic field tensor. Here it would also be unnatural to introduce curvature terms. They would cause a breakdown in charge conservation, in the sense of termination of electric and magnetic field lines at points where there is curvature but no charge. To maintain charge conservation, one omits curvature coupling when one translates Maxwell's equations (3.32) and (3.36) into curved spacetime:

$$F^{\alpha\beta}{}_{;\beta} = 4\pi J^{\alpha}, \qquad F_{\alpha\beta;\gamma} + F_{\beta\gamma;\alpha} + F_{\gamma\alpha;\beta} = 0.$$

Moreover, one continues to regard $F_{\mu\nu}$ as arising from a vector potential by the curved-spacetime translation of (3.54′)

$$F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu}.$$

These points granted, one can verify that the second of Maxwell's equations is automatically satisfied, and verify also that the first is satisfied if and only if

$$-A^{\alpha;\mu}{}_{\mu} + A^{\mu}{}_{;\mu}{}^{\alpha} + R^{\alpha}{}_{\mu}A^{\mu} = 4\pi J^{\alpha}.$$

(See §22.4 for fuller discussion and derivation.)

(3) *Transport law for Earth's angular-momentum vector.* If the Earth were in flat spacetime, like any other isolated body it would parallel-transport its angular-momentum vector $\boldsymbol{S}$ along the straight world line of its center of mass, $\nabla_{\boldsymbol{u}}\boldsymbol{S} = 0$ ("conservation of angular momentum"). When translating this transport law into curved spacetime (where the Earth actually resides!), can one ignore curvature coupling? No! Spacetime curvatures due to the moon and sun produce tidal gravitational forces in the Earth; and because the Earth has an equatorial bulge, the tidal forces produce a nonzero net torque about the Earth's center of mass. (In Newtonian language: the piece of bulge nearest the Moon gets pulled with greater force, and hence greater torque, than the piece of bulge farthest from the Moon.) Thus, in curved spacetime one expects a transport law of the form

$$\nabla_{\boldsymbol{u}}\boldsymbol{S} = (\text{Riemann tensor}) \times (\text{Earth's quadrupole moment}).$$

This curvature-coupling torque produces a precession of the Earth's rotation axis through a full circle in the plane of the ecliptic once every 26,000 years ("general precession"; "precession of the equinoxes"; discovered by Hipparchus about 150 B.C.). The precise form of the curvature-coupling term is derived in exercise 16.4.

Exercise 16.4. PRECESSION OF THE EQUINOXES

(a) Show that the transport law for the Earth's intrinsic angular momentum vector S^α in curved spacetime is

$$\frac{DS^\alpha}{D\tau} = \epsilon^{\alpha\beta\gamma\delta} I\!\!\!-_{\beta\mu} R^\mu{}_{\nu\gamma\zeta} u_\delta u^\nu u^\zeta. \tag{16.8}$$

Here $d/d\tau = \boldsymbol{u}$ is 4-velocity along the Earth's world line; $I\!\!\!-_{\beta\mu}$ is the Earth's "reduced quadrupole moment" (trace-free part of second moment of mass distribution), defined in the Earth's local Lorentz frame by

$$I\!\!\!-_{\hat{0}\hat{0}} = I\!\!\!-_{\hat{0}\hat{j}} = 0, \qquad I\!\!\!-_{\hat{j}\hat{k}} = \int \rho(x^{\hat{j}} x^{\hat{k}} - \frac{1}{3} \hat{r}^2 \delta_{jk})\, d^3\hat{x}; \tag{16.9}$$

and $R^\mu{}_{\nu\gamma\zeta}$ is the Riemann curvature produced at the Earth's location by the moon, sun, and planets. [*Hint:* Derive this result in the Earth's local Lorentz frame, ignoring the spacetime curvature due to the Earth. (In this essentially Newtonian situation, curvature components $R^{\hat{j}}{}_{\hat{0}\hat{k}\hat{0}}$ due to the Earth, sun, moon, and planets superpose linearly; "gravity too weak to be nonlinear"). Integrate up the torque produced about the Earth's center of mass by tidal gravitational forces ("geodesic deviation"):

$$\begin{pmatrix}\text{acceleration at } x^{\hat{j}}, \text{ relative to center of mass } (x^{\hat{j}} = 0),\\ \text{produced by tidal gravitational forces but counterbalanced}\\ \text{in part by Earth's internal stresses}\end{pmatrix}$$

$$= \left(\frac{d^2 x^{\hat{k}}}{d\hat{t}^2}\right)_{\substack{\text{geodesic}\\ \text{deviation}}} = -R^{\hat{k}}{}_{\hat{0}\hat{l}\hat{0}} x^{\hat{l}} \text{ [see equation (1.8′)]};$$

$$\begin{pmatrix}\text{force per unit volume due to this}\\ \text{acceleration, relative to center}\\ \text{of mass}\end{pmatrix}^{\hat{k}} = \underbrace{\rho}_{\text{mass density}} \frac{d^2 x^{\hat{k}}}{d\hat{t}^2} = -\rho R^{\hat{k}}{}_{\hat{0}\hat{l}\hat{0}} x^{\hat{l}};$$

$$\begin{pmatrix}\text{torque per unit volume relative}\\ \text{to center of mass}\end{pmatrix}_{\hat{i}} = \epsilon_{\hat{0}\hat{i}\hat{j}\hat{k}} x^{\hat{j}}(-\rho R^{\hat{k}}{}_{\hat{0}\hat{l}\hat{0}} x^{\hat{l}});$$

$$\begin{pmatrix}\text{total torque about center}\\ \text{of mass}\end{pmatrix}_{\hat{i}} = \int [\epsilon_{\hat{0}\hat{i}\hat{j}\hat{k}} x^{\hat{j}}(-\rho R^{\hat{k}}{}_{\hat{0}\hat{l}\hat{0}} x^{\hat{l}})]\, d^3\hat{x}.$$

Put this expression into a form involving $I\!\!\!-_{\hat{j}\hat{l}}$, equate it to $dS_{\hat{i}}/d\tau$, and then reexpress it in frame-independent, component notation. The result should be equation (16.8).]

(b) Rewrite equation (16.8) in the Earth's local Lorentz frame, using the equation

$$R^{\hat{j}}{}_{\hat{0}\hat{k}\hat{0}} = \partial^2\Phi/\partial x^{\hat{j}}\,\partial x^{\hat{k}}$$

for the components of ***Riemann*** in terms of the Newtonian gravitational potential. (Newtonian approximation to Einstein theory. Track-2 readers have met this equation in Chapter 12; track-one readers will meet it in §17.4.)

(c) Calculate $dS^{\hat{j}}/d\hat{t}$ using Newton's theory of gravity from the beginning. The answer should be identical to that obtained in part (b) using Einstein's theory.

(d) Idealizing the moon and sun as point masses, calculate the long-term effect of the spacetime curvatures that they produce upon the Earth's rotation axis. Use the result of part (b), together with moderately accurate numerical values for the relevant solar-system parameters. [*Answer:* The Earth's rotation axis precesses relative to the axes of its local Lorentz frame ("precession of the equinoxes"; "general precession"); the precession period is 26,000 years. The details of the calculation will be found in any textbook on celestial mechanics.]

§16.4. THE RODS AND CLOCKS USED TO MEASURE SPACE AND TIME INTERVALS

Turn attention now from the laws of physics in the presence of gravity to the nature of the rods and clocks that must be used for measuring the length and time intervals appearing in those laws.

One need not—and indeed must not!—postulate that proper length s is measured by a certain type of rod (e.g., platinum meter stick), or that proper time τ is measured by a certain type of clock (e.g., hydrogen-maser clock). Rather, one must ask the laws of physics themselves what types of rods and clocks will do the job. Put differently, one *defines an "ideal" rod or clock* to be one which measures proper length as given by $ds = (g_{\alpha\beta}\, dx^\alpha\, dx^\beta)^{1/2}$ or proper time as given by $d\tau = (-g_{\alpha\beta}\, dx^\alpha\, dx^\beta)^{1/2}$ (the kind of clock to which one was led by physical arguments in §1.5). One must then determine the accuracy to which a given rod or clock is ideal under given circumstances by using the laws of physics to analyze its behavior.

"Ideal" rods and clocks defined

How ideal are real clocks?

As an obvious example, consider a pendulum clock. If it is placed at rest on the Earth's surface, if it is tiny enough that redshift effects from one end to the other and time dilation effects due to its swinging velocity are negligible, and if the accuracy one demands is small enough that time variations in the local gravitational acceleration due to Earth tides can be ignored, then the laws of physics report (Box 16.2) that the pendulum clock is "ideal." However, in any other context (e.g., on a rocket journey to the moon), a pendulum clock should be far from ideal. Wildly changing accelerations, or no acceleration at all, will make it worthless!

(1) pendulum clocks

Of greater interest are atomic and nuclear clocks of various sorts. Such a clock is analyzed most easily if it is freely falling. One can then study it in its local Lorentz rest frame, using the standard equations of quantum theory; and, of course, one will find that it measures proper time to within the precision ($\Delta t/t \sim 10^{-9}$ to 10^{-14}) of the technology used in its construction. However, one rarely permits his atomic clock to fall freely. (The impact with the Earth's surface can be expensive!) Nevertheless, even when accelerated at "1 g" $= 980\ \text{cm/sec}^2$ on the Earth's surface, and even when accelerated at "2 g" in an airliner trying to avoid a midair collision (Box 16.3), an atomic clock—if built solidly—will still measure proper time $d\tau = (-g_{\alpha\beta}\, dx^\alpha\, dx^\beta)^{1/2}$ along its world line to nearly the same accuracy as if it were freely falling. To discover this one can perform an experiment. Alternatively, one can analyze the clock in its own "proper reference frame" (§13.6), with Fermi-Walker-transported basis vectors, using the standard local Lorentz laws of quantum mechanics as adapted to accelerated frames (local Lorentz laws plus an "inertial force," which can be treated as due to a potential with a uniform gradient.

(2) atomic clocks

Of course, any clock has a "breaking point," beyond which it will cease to function properly (Box 16.3). But that breaking point depends entirely on the construction of the clock—and not at all on any "universal influence of acceleration on the march of time." Velocity produces a universal time dilation; acceleration does not.

The aging of the human body is governed by the same electromagnetic and quantum-mechanical laws as govern the periodicities and level transitions in atoms and molecules. Consequently, aging, like atomic processes, is tied to proper time

(3) human clocks

(continued on page 396)

Box 16.2 PROOF THAT A PENDULUM CLOCK AT REST ON THE EARTH'S SURFACE IS IDEAL

That is, a proof that it measures the interval $d\tau = (-g_{\alpha\beta}\, dx^\alpha\, dx^\beta)^{1/2}$.

A. Constraint on the Pendulum

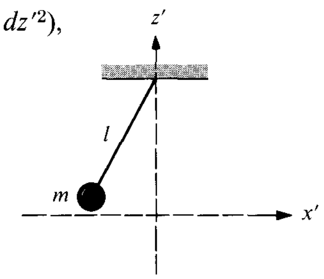
It must be so small that it cannot couple to the spacetime curvature—i.e., so small that the Earth's gravitational field looks uniform in its neighborhood—and that the velocity of its ball is totally negligible compared to the speed of light.

B. Coordinate System and Metric

(1) General coordinate system: because the Earth's field is nearly Newtonian, one can introduce the coordinates of "linearized theory" (§18.4; one must take this on faith until one reaches that point) in which

$$ds^2 = -(1 + 2\Phi)\, dt'^2 + (1 - 2\Phi)(dx'^2 + dy'^2 + dz'^2),$$

where Φ is the Newtonian potential.



(2) Put the origin of coordinates at the pendulum's equilibrium position, and orient the x',z'-plane so the pendulum swings in it.

(3) Renormalize the coordinates so they measure proper length and proper time at the equilibrium position

$$t = [1 + 2\Phi(0)]^{1/2} t', \qquad x^j = [1 - 2\Phi(0)]^{1/2} x^{j'}.$$

Then near the pendulum (inhomogeneities in the field neglected!)

$$\Phi = \Phi(0) + gz, \qquad g = \text{“acceleration of gravity,”} \tag{1}$$

$$ds^2 = -(1 + 2gz)\,dt^2 + (1 - 2gz)(dx^2 + dy^2 + dz^2). \tag{2}$$

C. Analysis of Pendulum Motion

(1) Put the total mass m of the pendulum in its ball (negligible mass in its rod). Let its rod have proper length l.

(2) Calculate the 4-acceleration $\boldsymbol{a} = \boldsymbol{\nabla_u u}$ of the pendulum's ball in terms of d^2x^α/dt^2, using the velocity condition $v <<< 1$ and $dt/d\tau \approx 1$:

$$\begin{aligned} a^x &= d^2x/d\tau^2 + \Gamma^x{}_{00}(dt/d\tau)^2 = d^2x/dt^2 + \Gamma^x{}_{00} = d^2x/dt^2, \\ a^z &= d^2z/d\tau^2 + \Gamma^z{}_{00}(dt/d\tau)^2 = d^2z/dt^2 + \Gamma^z{}_{00} = d^2z/dt^2 + g. \end{aligned} \tag{3}$$

(3) This 4-acceleration must be produced by the forces in the rod, and must be directed up the rod so that (for $x \ll l$ so $g \gg d^2z/dt^2$)

$$d^2x/dt^2 = a^x = -(x/l)a^z = -(g/l)x. \tag{4}$$

(4) Solve this differential equation to obtain

$$x = x_0 \cos\left(t\sqrt{g/l}\right). \tag{5}$$

(5) Thus conclude that the pendulum is periodic in t, which is proper time at the ball's equilibrium position (see equation 2). This means that *the pendulum is an ideal clock when it is at rest on the Earth's surface.*

Note: The above analysis ignores the Earth's rotation; for an alternative analysis including rotation, one can perform a similar calculation at the origin of the pendulum's "proper reference frame" [§13.6; line element (13.71)]. The answer is the same; but now "g" is a superposition of the "gravitational acceleration," and the "centrifugal acceleration produced by Earth's rotation."

Box 16.3 RESPONSE OF CLOCKS TO ACCELERATION AND TO TIDAL GRAVITATIONAL FORCES

Consider an atomic clock with frequency stabilized by some atomic or molecular process—for example, fixed by the "umbrella vibrations" of ammonia molecules [see Feynman *et. al.* (1964)]. When subjected to sufficiently strong accelerations or tidal forces, such a clock will cease to measure proper time with its normal precision. Two types of effects could lead to such departures from "ideality":

A. Influence of the acceleration or tidal force on the atomic process that provides the frequency stability. Example: If tidal forces are significant over distances of a few angstroms (e.g., near a spacetime "singularity" terminating gravitational collapse), then they can and will deform an ammonia molecule and destroy the regularity of its umbrella vibrations, thereby making useless *any* ammonia atomic clock, no matter how constructed. Similarly, if an ammonia molecule is subjected to accelerations of magnitude comparable to its internal atomic accelerations ($a \sim 10^{12}$"g" $\sim 10^{15}$ cm/sec^2), which change in times of the order of the "umbrella" vibration period, then it must cease to vibrate regularly, and any clock based on its vibrations must fail. Such limits of principle on the ideality of a clock will vary from one atomic process to another. However, they are far from being a limiting factor on clock construction in 1973. Much more important today is:

B. Influence of the acceleration or tidal force on the macroscopic structure of the clock—a structure dictated by current technology. The crystal oscillator, which produces the periodic signal output, must be locked to the regulating atomic process in some way. The lock will be disturbed by moderate accelerations. The toughest task for the manufacturer of aircraft clocks is to guarantee that precise locking will be maintained, even when the aircraft is maneuvering desperately to avoid collision with another aircraft or with a missile. In 1972 a solidly built rubidium clock will maintain its lock, with no apparent degradation of stability

$$[\Delta t/t \sim 10^{-12}(1 \text{ sec}/t)^{1/2} \;\text{ for }\; 1 \text{ sec} \lesssim t \lesssim 10^3 \text{ sec}]$$

under steady-state accelerations up to 50 "g" or more. But, because of the finite bandwidth of the lock loop (typically $\Delta\nu \sim 20$ to 50 Hz), sudden changes in acceleration will temporarily break the lock, degrading the clock stability to that of the unlocked crystal oscillator—for which an acceleration a produces a change in frequency of about (a/1 "g") $\times 10^{-9}$. But the lock to the rubidium standard is restored quickly ($\delta t \sim 1/\Delta\nu$), bringing the clock back to its normal highly stable performance.*

Tidal forces are so small in the solar system that the clock manufacturer can ignore them. However, a 1973 atomic clock, subjected to the tidal accelerations near a spacetime singularity, should break the "lock" to its atomic process long before the tidal forces become strong enough to influence the atomic process itself.

*For this information on the response of rubidium clocks to acceleration, we thank H. P. Stratemeyer of General Radio Company, Concord, Massachusetts.

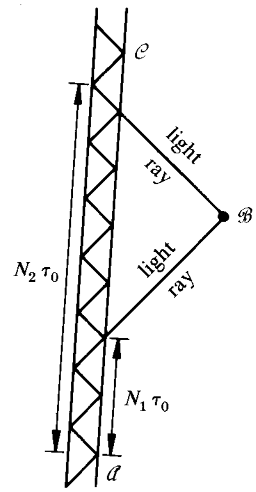
as governed by the metric—though, of course, it is also tied to other things, such as cigarette smoking.

Ideal rods and clocks constructed from geodesic world lines

In principle, one can build ideal rods and clocks from the geodesic world lines of freely falling test particles and photons. (See Box 16.4.) In other words, spacetime has its own rods and clocks built into itself, even when matter and nongravitational fields are absent!

Box 16.4 IDEAL RODS AND CLOCKS BUILT FROM GEODESIC WORLD LINES*

The Standard Interval. A specific timelike interval—the interval between two particular neighboring events $\mathcal{A}$ and $\mathcal{B}$—is chosen as the standard interval, and is assigned unit length. It is used to calibrate a huge set of geodesic clocks that pass through $\mathcal{A}$.

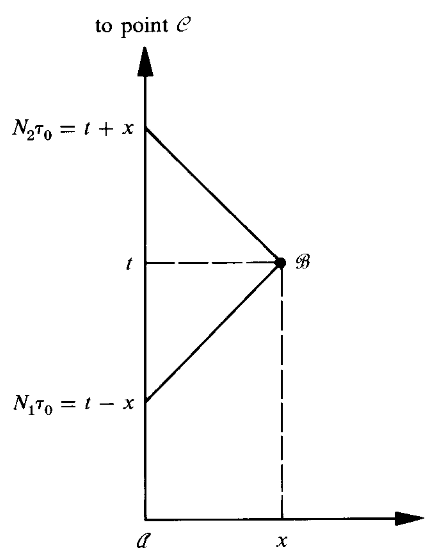


Each *geodesic clock* is constructed and calibrated as follows:

(1) A timelike geodesic $\mathcal{A}\mathcal{C}$ (path of freely falling particle) passes through $\mathcal{A}$.
(2) A neighboring world line, everywhere parallel to $\mathcal{A}\mathcal{C}$ (and thus not a geodesic), is constructed by the method of Schild's ladder (Box 10.2), which relies only on geodesics.
(3) Light rays (null geodesics) bounce back and forth between these parallel world lines; each round trip constitutes one "tick."
(4) The proper time lapse, τ_0, between ticks is related to the interval $\mathcal{A}\mathcal{B}$ by

$$-1 \equiv (\mathcal{A}\mathcal{B})^2 = -(N_1\tau_0)(N_2\tau_0),$$

where N_1 and N_2 are the number of ticks between the events shown in the diagrams. [*Proof:* see diagram at right.]



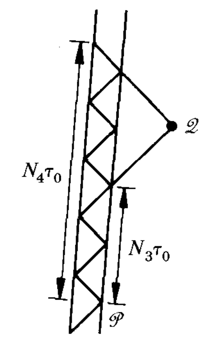
In local Lorentz rest frame of geodesic clock:

$$\begin{aligned}(N_1\tau_0)(N_2\tau_0) &= (t - x)(t + x)\\ &= t^2 - x^2 = -(\mathcal{A}\mathcal{B})^2\end{aligned}$$

Spacetime is filled with such geodesic clocks. Those that pass through $\mathcal{A}$ are calibrated as above against the standard interval $\mathcal{A}\mathcal{B}$, and are used subsequently to calibrate all other clocks they meet.

*Based on Marzke and Wheeler (1964).

Box 16.4 (*continued*)



Any interval $\mathcal{PQ}$ along the world line of a geodesic clock can be measured by the same method as was used in calibration. The interval $\mathcal{PQ}$ can be timelike, spacelike, or null; its squared length in all three cases will be

$$(\mathcal{PQ})^2 = -(N_3\tau_0)(N_4\tau_0)$$

To achieve a precision of measurement good to one part in N, where N is some large number, take two precautions:

(1) Demand that the intervals $\mathcal{AB}$ and $\mathcal{PQ}$ be sufficiently small compared to the scale of curvature of spacetime; or specifically,

$$R^{(AB)}(\mathcal{AB})^2 \ll 1/N$$

and

$$R^{(PQ)}(\mathcal{PQ})^2 \ll 1/N,$$

where $R^{(AB)}$ and $R^{(PQ)}$ are the largest relevant components of the curvature tensor in the two regions in question.

(2) Demand that the time scale, τ_0, of the geodesic clocks employed be small compared to $\mathcal{AB}$ and $\mathcal{PQ}$ individually; thus,

$$\tau_0 \ll \mathcal{AB}/N,$$
$$\tau_0 \ll \mathcal{PQ}/N.$$

The Einstein principle that spacetime is described by Riemannian geometry exposes itself to destruction by a "thousand" tests. Thus, from the fiducial interval, $\mathcal{AB}$, to the interval under measurement, $\mathcal{PQ}$, there are a "score" of routes of intercomparison, all of which must give the same value for the ratio $\mathcal{PQ}/\mathcal{AB}$. Moreover, one can easily select out "fifty" intervals $\mathcal{PQ}$ to which the same kind of test can be applied. Such tests are not all items for the future.

Some 5×10^9 years ago, electrons arrived by different routes at a common location, a given atom of iron in the core of the earth. This iron atom does not collapse. The Pauli principle of exclusion keeps the electrons from all falling into the K-orbit. The Pauli principle would not apply if the electrons were not identical or nearly so. From this circumstance it would appear possible to draw an important conclusion (Marzke and Wheeler). With each electron is associated a standard length, its Compton wavelength, $\hbar/mc$. If these lengths had started different, or changed by different amounts along the different routes, and if the resulting difference in properties were as great as one part in

$$\sim(5 \times 10^9 \text{ yr}) \times (3 \times 10^7 \text{ sec/yr}) \times (5 \times 10^{18} \text{ rev/sec}) \sim 10^{36},$$

by now this difference would have shown up, the varied electrons would have fallen into the K-orbit, and the earth would have collapsed, contrary to observation.

The Marzke-Wheeler construction expresses an arbitrary small interval $\mathcal{P}\mathcal{Q}$, anywhere in spacetime, in terms of the fiducial interval $\mathcal{A}\mathcal{B}$, an interval which itself may be taken for definiteness to be the "geometrodynamic standard centimeter" of §1.5. This construction thus gives a vivid meaning to the idea of Riemannian geometry.

The M-W construction makes no appeal whatsoever to rods and clocks of atomic constitution. This circumstance is significant for the following reasons. The length of the usual platinum meter stick is some multiple, $N_1(\hbar^2/me^2)$, of the Bohr atomic radius. Similarly, the wavelength of the Kr^{86} line is some multiple, $N_2(\hbar c/e^2)(\hbar^2/me^2)$, of a second basic length that depends on the atomic constants in quite a different way. Thus, if there is any change with time in the dimensionless ratio $\hbar c/e^2 = 137.038$, one or the other or both of these atomic standards of length must get out of kilter with the geometrodynamic standard centimeter. In this case, general relativity says, "Stick to the geometrodynamic standard centimeter."

Hermann Weyl at first thought that one could carry out the comparison of lengths by light rays alone, but H. A. Lorentz pointed out that one can dispense with the geodesics neither of test particles nor of light rays in the measurement process, the construction for which, however, neither Weyl nor Lorentz supplied [literature in Marzke and Wheeler (1964)]. Ehlers, Pirani, and Schild (1972) have given a deeper analysis of the separate parts played in the measurement process by the affine connection, by the conformal part of the metric, and by the full metric.

§16.5. THE MEASUREMENT OF THE GRAVITATIONAL FIELD

"I know how to measure the electromagnetic field using test charges; what is the analogous procedure for measuring the gravitational field?" This question has, at the same time, many answers and none.

The many faces of gravity, and how one measures them

It has no answers because nowhere has a precise definition of the term "gravitational field" been given—nor will one be given. Many different mathematical entities are associated with gravitation: the metric, the Riemann curvature tensor, the Ricci curvature tensor, the curvature scalar, the covariant derivative, the connection coefficients, etc. Each of these plays an important role in gravitation theory, and none is so much more central than the others that it deserves the name "gravitational field." Thus it is that throughout this book the terms "gravitational field" and "gravity" refer in a vague, collective sort of way to all of these entities. Another, equivalent term used for them is the "geometry of spacetime."

To "measure the gravitational field," then, means to "explore experimentally various properties of the spacetime geometry." One makes different kinds of measurements, depending on which geometric property of spacetime one is interested in. However, all such measurements must involve a scrutiny of the effects of the spacetime geometry (i.e., of gravity) on particles, on matter, or on nongravitational fields.

For example, to "measure" the metric near a given event, one typically lays out a latticework of rods and clocks (local orthonormal frame, small enough that curvature effects are negligible), and uses it to determine the interval between neighboring events. To measure the Riemann curvature tensor near an event, one typically studies the geodesic deviation (relative accelerations) that curvature produces between the world lines of a variety of neighboring test particles; alternatively, one makes measurements with a "gravity gradiometer" (Box 16.5) if the curvature is static or slowly varying; or with a gravitational wave antenna (Chapter 37) if the curvature fluctuates rapidly. To study the large-scale curvature of spacetime, one examines large-scale effects of gravity, such as the orbits of planets and satellites, or the bending of light by the sun's gravitational field.

But whatever aspect of gravity one measures, and however one measures it, one is studying the geometry of spacetime.

EXERCISE

Exercise 16.5. GRAVITY GRADIOMETER

The gravity gradiometer of Box 16.5 moves through curved spacetime along an accelerated world line. Calculate the amplitude and phase of oscillation of one arm of the gradiometer relative to the other. [Hint: Perform the calculation in the gradiometer's "proper reference frame" (§13.6), with Fermi-Walker-transported basis vectors. Use, as the equation for the relative angular acceleration of the two arms,

$$2m\ell^2(\ddot{\alpha} + \dot{\alpha}/\tau_0 + {\omega_0}^2\alpha) = \begin{pmatrix}\text{Driving torque produced by}\\ \text{Riemann curvature}\end{pmatrix},$$

where

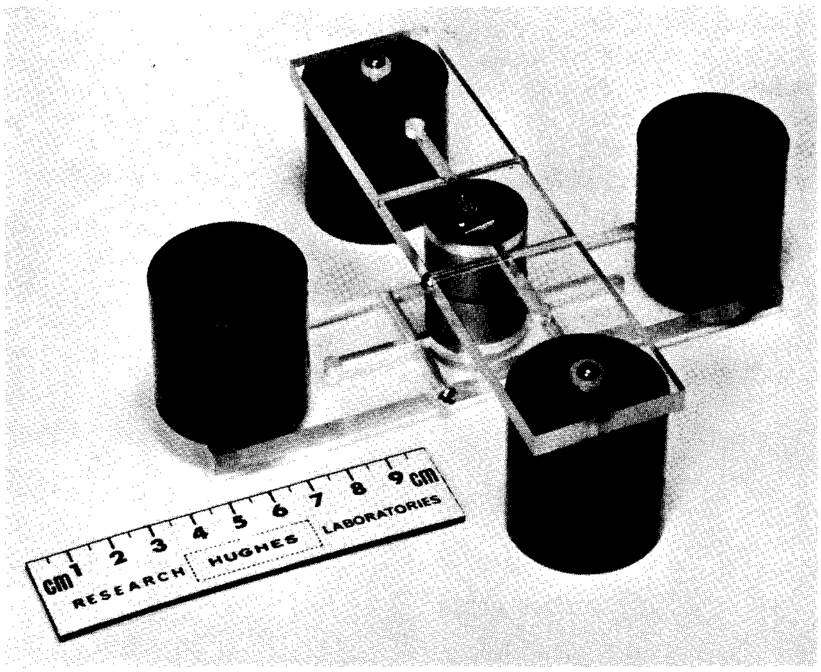
$$\begin{aligned}
2m\ell^2 &= (\text{moment of inertia of one arm}),\\
\alpha &= (\text{angular displacement of one arm from equilibrium}),\\
\frac{\pi}{2} + 2\alpha &= (\text{angular separation of the two arms}),\\
2m\ell^2{\omega_0}^2 &= (\text{torsional spring constant}),\\
\omega_0 &= (\text{angular frequency of free vibrations}),\\
\tau_0 &= (\text{decay time for free vibrations to damp out due to}\\
&\qquad \text{internal frictional forces}).
\end{aligned}$$

If $\boldsymbol{\xi}$ is the vector from the center of mass of the gradiometer to mass 1, then one has

$$\begin{pmatrix}\text{curvature-produced}\\ \text{acceleration of mass 1}\\ \text{relative to center of}\\ \text{gradiometer}\end{pmatrix}_{\hat{k}} = \left(\frac{D^2\xi_{\hat{k}}}{d\tau^2}\right)_{\substack{\text{geodesic}\\ \text{deviation}}} = -R_{\hat{k}\hat{0}\hat{l}\hat{0}}\xi_{\hat{l}};$$

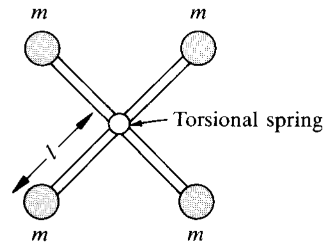
(*continued on page 403*)

Box 16.5 GRAVITY GRADIOMETER FOR MEASURING THE RIEMANN CURVATURE OF SPACETIME



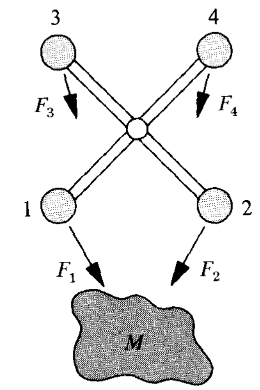
This gravity gradiometer was designed and built by Robert M. Forward and his colleagues at Hughes Research Laboratories, Malibu, California. It measures the Riemann curvature of spacetime produced by nearby masses. By flying a more advanced version of such a gradiometer in an airplane above the Earth's surface, one should be able to measure subsurface mass variations due to varying geological structure. In an Earth-orbiting satellite, such a gradiometer could measure the gravitational multipole moments of the Earth. Technical details of the gradiometer are spelled out in the papers of Forward (1972), and Bell, Forward, and Williams (1970). The principles of its operation are outlined below.

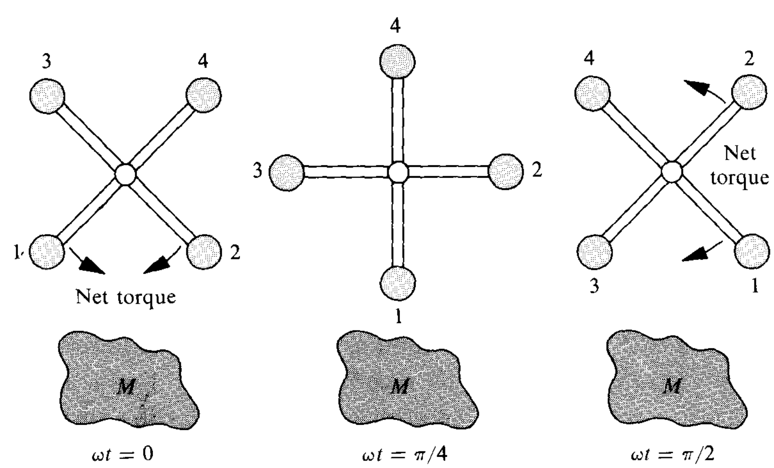
The gradiometer consists of two orthogonal arms with masses m on their ends, connected at their centers by a torsional spring. When the arms are twisted out of orthogonal alignment, they oscillate. A piezoelectric strain transducer is used to measure the oscillation amplitude.



Box 16.5 (*continued*)

When placed near an external mass, M, the gradiometer experiences a torque: because of the gradient in the gravitational field of M (i.e., because of the spacetime curvature produced by M), the Newtonian forces F_1 and F_2 are greater than F_3 and F_4; so a net torque pulls masses 1 and 2 toward each other, and 3 and 4 toward each other. [*Note:* the forces F_1, F_2, F_3, F_4 depend on whether the gradiometer is in free fall (geodesic motion; $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0$) or is moving on an accelerated world line. But the net torque is unaffected by acceleration; acceleration produces equal Newtonian forces on all four masses, with zero net torque.]





When in operation the gradiometer rotates with angular velocity ω about its center. As it rotates, the torques on its arms oscillate:

at $\omega t = 0$ net torque pushes 1 and 2 toward each other;
at $\omega t = \pi/4$ net torque is zero;
at $\omega t = \pi/2$ net torque pushes 1 and 2 away from each other.

The angular frequency of the oscillating torque is 2ω. If 2ω is set equal to $\omega_0 \equiv$ (natural oscillation frequency of the arms), the oscillating torque drives the arms into resonant oscillation. The resulting oscillation amplitude, in the 1970 prototype of the gradiometer, was easily detectable for gravity gradients (Riemann curvatures) of magnitude

$$\gtrsim 0.0002\left[\frac{2(\text{mass of earth})}{(\text{radius of earth})^3}\right]$$
$$\sim 1 \times 10^{-30}\ \text{cm}^{-2} \sim .01\ \text{g/cm}^3$$

$\sim$ [Riemann curvature produced by a two-kilometer high mountain, idealized as a two-kilometer high cube, at a distance of 15 kilometers. (Neglected in this idealization are isostacy and any lowering of density of Earth's crust in regions of mountain uplift.)]

For a mathematical analysis of the gradiometer, see exercise 16.5.

$$\begin{pmatrix}\text{torque acting on mass 1} \\ \text{relative to center of} \\ \text{gradiometer}\end{pmatrix}_{\hat{i}} = \epsilon_{\hat{i}\hat{j}\hat{k}}\xi_{\hat{j}}(-mR_{\hat{k}\hat{0}\hat{l}\hat{0}}\xi_{\hat{l}}).$$

The torque on mass 4 is identical to this (replace $\boldsymbol{\xi}$ by $-\boldsymbol{\xi}$), so the total torque on arm 1-4 is twice this. The components $R_{\hat{k}\hat{0}\hat{l}\hat{0}}$ of ***Riemann*** can be regarded as components of a 3×3 symmetric matrix. By appropriate orientation of the reference frame's spatial axes (orientation along "principal axes" of $R_{\hat{k}\hat{0}\hat{l}\hat{0}}$), one can make $R_{\hat{k}\hat{0}\hat{l}\hat{0}}$ diagonal at some initial moment of time

$$R_{\hat{x}\hat{0}\hat{x}\hat{0}} \neq 0,\; R_{\hat{y}\hat{0}\hat{y}\hat{0}} \neq 0,\; R_{\hat{z}\hat{0}\hat{z}\hat{0}} \neq 0, \text{ all others vanish.}$$

Assume that ***Riemann*** changes sufficiently slowly along the gradiometer's world line that throughout the experiment $R_{\hat{j}\hat{0}\hat{k}\hat{0}}$ remains diagonal and constant. For simplicity, place the gradiometer in the $\hat{x}$, $\hat{y}$-plane, so it rotates about the $\hat{z}$ axis with angular velocity $\omega \approx \frac{1}{2}\omega_0$:

$$\begin{pmatrix}\text{Angle of arm 1-4} \\ \text{relative to } \hat{x} \text{ axis}\end{pmatrix} = \omega t.$$

Show that the resultant equation of oscillation is

$$\ddot{\alpha} + \dot{\alpha}/\tau_0 + \omega_0{}^2\alpha = \frac{1}{2}(R_{\hat{x}\hat{0}\hat{x}\hat{0}} - R_{\hat{y}\hat{0}\hat{y}\hat{0}})\sin 2\omega t;$$

and that the steady-state oscillations are

$$\alpha = Im\left\{\frac{\frac{1}{2}(R_{\hat{x}\hat{0}\hat{x}\hat{0}} - R_{\hat{y}\hat{0}\hat{y}\hat{0}})}{2\omega_0(\omega_0 - 2\omega + i/2\tau_0)}e^{i2\omega t}\right\}.$$

Thus, for fixed ω (e.g., $2\omega = \omega_0$), by measuring the amplitude and phase of the oscillations, one can learn the magnitude and sign of $R_{\hat{x}\hat{0}\hat{x}\hat{0}} - R_{\hat{y}\hat{0}\hat{y}\hat{0}}$. The other differences, $R_{\hat{y}\hat{0}\hat{y}\hat{0}} - R_{\hat{z}\hat{0}\hat{z}\hat{0}}$ and $R_{\hat{z}\hat{0}\hat{z}\hat{0}} - R_{\hat{x}\hat{0}\hat{x}\hat{0}}$ can be measured by placing the gradiometer's rotation axis along the $\hat{x}$ and $\hat{y}$ axes, respectively.]

1

CHAPTER 17

HOW MASS-ENERGY GENERATES CURVATURE

The physical world is represented as a four-dimensional continuum. If in this I adopt a Riemannian metric, and look for the simplest laws which such a metric can satisfy, I arrive at the relativistic gravitation theory of empty space. If I adopt in this space a vector field, or the antisymmetrical tensor field derived from it, and if I look for the simplest laws which such a field can satisfy, I arrive at the Maxwell equations for free space. . . . at any given moment, out of all conceivable constructions, a single one has always proved itself absolutely superior to all the rest . . .

ALBERT EINSTEIN (1934, p. 18)

§17.1. AUTOMATIC CONSERVATION OF THE SOURCE AS THE CENTRAL IDEA IN THE FORMULATION OF THE FIELD EQUATION

This section derives the "Einstein field equation"

Turn now from the response of matter to geometry (motion of a neutral test particle on a geodesic; "comma-goes-to-semicolon rule" for the dynamics of matter and fields), and analyze the response of geometry to matter.

Mass is the source of gravity. The density of mass-energy as measured by any observer with 4-velocity $\boldsymbol{u}$ is

$$\rho = \boldsymbol{u} \cdot \boldsymbol{T} \cdot \boldsymbol{u} = u^{\alpha} T_{\alpha\beta} u^{\beta}. \tag{17.1}$$

Therefore the stress-energy tensor $\boldsymbol{T}$ is the frame-independent "geometric object" that must act as the source of gravity.

This source, this geometric object, is not an arbitrary symmetric tensor. It must have zero divergence

$$\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0, \tag{17.2}$$

because only so can the law of conservation of momentum-energy be upheld.

Equation describing how matter generates gravity must have form $\boldsymbol{G} = \kappa \boldsymbol{T}$, where $\boldsymbol{T}$ is stress-energy tensor

Place this source, $\boldsymbol{T}$, on the righthand side of the equation for the generation of gravity. On the lefthand side will stand a geometric object that characterizes gravity. That object, like $\boldsymbol{T}$, must be a symmetric, divergence-free tensor; and if it is to characterize gravity, it must be built out of the geometry of spacetime and nothing but that geometry. Give this object the name "Einstein tensor" and denote it by $\boldsymbol{G}$, so that the equation for the generation of gravity reads

$$\boldsymbol{G} = \underbrace{\kappa}_{\begin{bmatrix}\text{proportionality factor;}\\ \text{to be evaluated later}\end{bmatrix}} \boldsymbol{T}. \tag{17.3}$$

(Do not assume that $\boldsymbol{G}$ is the same Einstein tensor as was encountered in Chapters 8, 13, 14, and 15; that will be proved below!)

Properties that the tensor $\boldsymbol{G}$ must have

The vanishing of the divergence $\boldsymbol{\nabla} \cdot \boldsymbol{G}$ is not to be regarded as a consequence of $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$. Rather, the obedience of all matter and fields to the conservation law $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ is to be regarded (1) as a consequence of the way [equation (17.3)] they are wired into the geometry of spacetime, and therefore (2) as required and enforced by an *automatic* conservation law, or *identity*, that holds for any smooth Riemannian spacetime whatsoever, physical or not: $\boldsymbol{\nabla} \cdot \boldsymbol{G} \equiv 0$. (See Chapter 15 for a fuller discussion and §17.2 below for a fuller justification.) Accordingly, look for a symmetric tensor $\boldsymbol{G}$ that is an "automatically conserved measure of the curvature of spacetime" in the following sense:

(1) $\boldsymbol{G}$ vanishes when spacetime is flat.
(2) $\boldsymbol{G}$ is constructed from the Riemann curvature tensor and the metric, and from nothing else.
(3) $\boldsymbol{G}$ is distinguished from other tensors which can be built from ***Riemann*** and $\boldsymbol{g}$ by the demands (i) that it be linear in ***Riemann***, as befits any natural measure of curvature; (ii) that, like $\boldsymbol{T}$, it be symmetric and of second rank; and (iii) that it have an automatically vanishing divergence,

$$\boldsymbol{\nabla} \cdot \boldsymbol{G} \equiv 0. \tag{17.4}$$

Proof that $\boldsymbol{G}$ must be the Einstein curvature tensor of Chapter 8

Apart from a multiplicative constant, there is only one tensor (exercise 17.1) that satisfies these requirements of being an automatically conserved, second-rank tensor, linear in the curvature, and of vanishing when spacetime is flat. It is the Einstein curvature tensor, $\boldsymbol{G}$, expressed in Chapter 8 in terms of the Ricci curvature tensor:

$$\begin{aligned} R_{\mu\nu} &= R^{\alpha}{}_{\mu\alpha\nu}, \\ G_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \end{aligned} \tag{17.5}$$

This quantity was given vivid meaning in Chapter 15 as the "moment of rotation of the curvature" or, more simply, the "moment of rotation," constructed by taking the double-dual

$$\boldsymbol{G} = {}^*\boldsymbol{Riemann}^* \tag{17.6a}$$

of the Riemann curvature tensor, and then contracting this double dual,

$$G_{\mu\nu} = G^{\alpha}{}_{\mu\alpha\nu}. \tag{17.6b}$$

In Chapter 15 the vanishing of $\boldsymbol{\nabla} \cdot \boldsymbol{G}$ was shown to follow as a consequence of the elementary principle of topology that "the boundary of a boundary is zero."

Evaluation of κ (in $\boldsymbol{G} = \kappa\boldsymbol{T}$) by comparing with Newtonian theory of gravity

To evaluate the proportionality constant κ in the "Einstein field equation" $\boldsymbol{G} = \kappa\boldsymbol{T}$, one can compare with the well-tested Newtonian theory of gravity. To facilitate the comparison, examine the relative acceleration (geodesic deviation) of particles that fall down a pipe inserted into an idealized Earth of uniform density ρ (Figure 1.12). According to Newton, the relative acceleration is governed by the density; according to Einstein, it is governed by the Riemann curvature of spacetime. Direct comparison of the Newtonian and Einstein predictions using Newtonian coordinates (where $g_{\mu\nu} \approx \eta_{\mu\nu}$) reveals the relation

$$R_{00} \equiv R^{\alpha}{}_{0\alpha 0} = 4\pi\rho. \tag{17.7}$$

(See §1.7 for details of the derivation; see Chapter 12 for extensive discussion of Newtonian gravity using this equation.) When applied to the Earth's interior, the Einstein field equation $\boldsymbol{G} = \kappa\boldsymbol{T}$ must thus reduce to $R_{00} = 4\pi\rho$. In component form, the Einstein field equation reads

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \kappa T_{\mu\nu}.$$

Its trace reads

$$-R = R - 2R = \kappa T.$$

In consequence, it predicts

$$\begin{aligned} R_{00} &= \frac{1}{2} g_{00} R + \kappa T_{00} = \frac{1}{2}\kappa(2T_{00} - \underbrace{g_{00}}_{-1} T) \\ &= \frac{1}{2}\kappa[2T_{00} + (T^0{}_0 + T^j{}_j)] \\ &= \frac{1}{2}\kappa(T_{00} + T^j{}_j), \end{aligned}$$

which reduces to

$$R_{00} = \frac{1}{2}\kappa\rho \tag{17.8}$$

when one recalls that for the Earth—as for any nearly Newtonian system—the stresses T_{jk} are very small compared to the density of mass-energy $T_{00} = \rho$:

$$\frac{|T_{jk}|}{T_{00}} \sim \frac{\text{pressure}}{\text{density}} \sim \frac{dp}{d\rho} \sim (\text{velocity of sound})^2 \ll 1.$$

The equation $R_{00} = 4\pi\rho$ (derived by comparing relative accelerations in the Newton and Einstein theories) and the equation $R_{00} = \frac{1}{2}\kappa\rho$ (derived directly from the Einstein field equation) can agree only if the proportionality constant κ is 8π.

Thus, the Einstein field equation, describing the generation of curvature by mass-energy, must read

Result: "Einstein field equation" $\boldsymbol{G} = 8\pi\boldsymbol{T}$

$$\boldsymbol{G} = 8\pi\boldsymbol{T}. \tag{17.9}$$

The lefthand side ("curvature") has units cm^{-2}, since a curvature tensor is a linear machine into which one inserts a displacement (units: cm) and from which one gets a relative acceleration (units: $\text{cm/sec}^2 \sim \text{cm/cm}^2 \sim \text{cm}^{-1}$). The right-hand side also has dimensions cm^{-2}, since it is a linear machine into which one inserts 4-velocity (dimensionless) and from which one gets mass density [units: $g/\text{cm}^3 \sim \text{cm/cm}^3 \sim \text{cm}^{-2}$; recall from equation (1.12) and Box 1.8 that $1\text{g} = (1\text{g}) \times (G/c^2) = (1\text{g}) \times (0.742 \times 10^{-28}\ \text{cm/g}) = 0.742 \times 10^{-28}\ \text{cm}$].

This concludes the simplest derivation of Einstein's field equation that has come to hand, and establishes its correspondence with the Newtonian theory of gravity under Newtonian conditions. That correspondence had to be worked out to determine the factor $\kappa = 8\pi$ on the righthand side of (17.9). Apart from this factor, the central point in the derivation was the demand for, and the existence of, a unique tensorial measure of curvature $\boldsymbol{G}$ with an identically vanishing divergence.

EXERCISES

Exercise 17.1. UNIQUENESS OF THE EINSTEIN TENSOR

(a) Show that the most general second-rank, symmetric tensor constructable from ***Riemann*** and $\boldsymbol{g}$, and linear in ***Riemann***, is

$$\begin{aligned} &aR_{\alpha\beta} + bRg_{\alpha\beta} + \Lambda g_{\alpha\beta} \\ &= aR^{\mu}{}_{\alpha\mu\beta} + bR^{\mu\nu}{}_{\mu\nu}g_{\alpha\beta} + \Lambda g_{\alpha\beta}, \end{aligned} \tag{17.10}$$

where a, b, and Λ are constants.

(b) Show that this tensor has an automatically vanishing divergence if and only if $b = -\frac{1}{2}a$.

(c) Show that, in addition, this tensor vanishes in flat spacetime, if and only if $\Lambda = 0$—i.e., if and only if it is a multiple of the Einstein tensor $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta}$. (Do not bother to prove that $\boldsymbol{\nabla}\cdot\boldsymbol{G} \equiv 0$; assume it as a result from Chapter 13.)

Exercise 17.2. NO TENSOR CONSTRUCTABLE FROM FIRST DERIVATIVES OF METRIC

Show that there exists *no* tensor with components constructable from the ten metric coefficients $g_{\alpha\beta}$ and their 40 first derivatives $g_{\alpha\beta,\mu}$—except the metric tensor $\boldsymbol{g}$, and products of it with itself; e.g., $\boldsymbol{g} \otimes \boldsymbol{g}$. [*Hint:* Assume there exists some other such tensor, and examine its hypothesized components in a local inertial frame.]

Exercise 17.3. *RIEMANN* AS THE ONLY TENSOR CONSTRUCTABLE FROM, AND LINEAR IN SECOND DERIVATIVES OF METRIC

Show that (1) ***Riemann***, (2) $\boldsymbol{g}$, and (3) tensors (e.g., ***Ricci***) formed from ***Riemann*** and $\boldsymbol{g}$ but linear in ***Riemann***, are the only tensors that (a) are constructable from the ten $g_{\alpha\beta}$, the 40 $g_{\alpha\beta,\mu}$, and the 100 $g_{\alpha\beta,\mu\nu}$, and (b) are linear in the $g_{\alpha\beta,\mu\nu}$. [*Hint:* Assume there exists some other such tensor, and examine its hypothesized components in an orthonormal, Riemann-normal coordinate system. Use equations (11.30) to (11.32).]

Exercise 17.4. UNIQUENESS OF THE EINSTEIN TENSOR

(a) Show that the Einstein tensor, $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta}$, is the only second-rank, symmetric tensor that (1) has components constructable solely from $g_{\alpha\beta}, g_{\alpha\beta,\mu}, g_{\alpha\beta,\mu\nu}$; (2) has components linear in $g_{\alpha\beta,\mu\nu}$; (3) has an automatically vanishing divergence, $\boldsymbol{\nabla}\cdot\boldsymbol{G} = 0$; and (4) vanishes in flat spacetime. This provides added motivation for choosing the Einstein tensor as the left side of the field equation $\boldsymbol{G} = 8\pi\boldsymbol{T}$.

(b) Show that, when condition (4) is dropped, the most general tensor is $\boldsymbol{G} + \Lambda\boldsymbol{g}$, where Λ is a constant. (See §17.3 for the significance of this.)

§17.2. AUTOMATIC CONSERVATION OF THE SOURCE: A DYNAMIC NECESSITY

The answer $\boldsymbol{G} = 8\pi\boldsymbol{T}$ is now on hand; but what is the question? An equation has been derived that connects the Einstein-Cartan "moment of rotation" $\boldsymbol{G}$ with the stress-energy tensor $\boldsymbol{T}$, but what is the purpose for which one wants this equation in the first place? If geometry tells matter how to move, and matter tells geometry how to curve, does one not have in one's hands a Gordian knot? And how then can one ever untie it?

Einstein field equation governs the evolution of spacetime geometry

The story is no different in character for the dynamics of geometry than it is for other branches of dynamics. To predict the future, one must first specify, on an "initial" hypersurface of "simultaneity," the position and velocity of every particle, and the amplitude and time-rate of change of every field that obeys a second-order wave equation. One can then evolve the particles and fields forward in time by means of their dynamic equations. Similarly, one must give information about the geometry and its first time-rate of change on the "initial" hypersurface if the Einstein field equation is to be able to predict completely and deterministically the future time-development of the entire system, particles plus fields plus geometry. (See Chapter 21 for details.)

If a prediction is to be made of the geometry, how much information has to be supplied for this purpose? The geometry of spacetime is described by the metric

$$ds^2 = g_{\alpha\beta}(\mathscr{P})\,dx^\alpha\,dx^\beta;$$

that is, by the ten functions $g_{\alpha\beta}$ of location $\mathscr{P}$ in spacetime. It might then seem that ten functions must be predicted; and, if so, that one would need for the task ten equations. Not so. Introduce a new set of coordinates $x^{\bar{\mu}}$ by way of the coordinate transformations

$$x^{\alpha} = x^{\alpha}(x^{\bar{\mu}}),$$

and find the same spacetime geometry, with all the same bumps, rills, and waves, described by an entirely new set of metric coefficients $g_{\bar{\alpha}\bar{\beta}}(\mathscr{P})$.

It would transgress the power as well as the duty of Einstein's "geometrodynamic law" $\boldsymbol{G} = 8\pi\boldsymbol{T}$ if, out of the appropriate data on the "initial-value hypersurface," it were to provide a way to calculate, on out into the future, values for all ten functions $g_{\alpha\beta}(\mathscr{P})$. To predict all ten functions would presuppose a choice of the coordinates; and to make a choice among coordinate systems is exactly what the geometrodynamic law cannot and must not have the power to do. That choice resides of necessity in the man who studies the geometry, not in the Nature that makes the geometry. The geometry in and by itself, like an automobile fender in and by itself, is free of coordinates. The coordinates are the work of man.

$\boldsymbol{G} = 8\pi\boldsymbol{T}$ must determine only six metric components; the other four are adjustable by changes of coordinates

It follows that the ten components $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ of the field equation must not determine completely and uniquely all ten components $g_{\mu\nu}$ of the metric. On the contrary, $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ must place only six independent constraints on the ten $g_{\mu\nu}(\mathscr{P})$, leaving four arbitrary functions to be adjusted by man's specialization of the four coordinate functions $x^{\alpha}(\mathscr{P})$.

How can this be so? How can the ten equations $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ be in reality only six? Answer: by virtue of the "automatic conservation of the source." More specifically, the identity $G^{\alpha\beta}{}_{;\beta} \equiv 0$ guarantees that the ten equations $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ contain the four "conservation laws" $T^{\alpha\beta}{}_{;\beta} = 0$. These four conservation laws—along with other equations—govern the evolution of the source. They do not constrain in any way the evolution of the geometry. The geometry is constrained only by the six remaining, independent equations in $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$.

$\boldsymbol{G} = 8\pi\boldsymbol{T}$ leaves four components of metric free because it satisfies four identities $0 \equiv \boldsymbol{\nabla}\cdot\boldsymbol{G} = 8\pi\boldsymbol{\nabla}\cdot\boldsymbol{T}$ ("automatic conservation of source")

When viewed in this way, the "automatic conservation of the source" is not merely a philosophically attractive principle. It is, in fact, an absolute dynamic necessity. Without "automatic conservation of the source," the ten $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ would place ten constraints on the ten $g_{\alpha\beta}$, thus fixing the coordinate system as well as the geometry. With "automatic conservation," the ten $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ place four constraints (local conservation of energy and momentum) on the source, and six constraints on the ten $g_{\alpha\beta}$, leaving four of the $g_{\alpha\beta}$ to be adjusted by adjustment of the coordinate system.

§17.3. COSMOLOGICAL CONSTANT

Einstein's motivation for introducing a cosmological constant

In 1915, when Einstein developed his general relativity theory, the permanence of the universe was a fixed item of belief in Western philosophy. "The heavens endure from everlasting to everlasting." Thus, it disturbed Einstein greatly to discover (Chapter 27) that his geometrodynamic law $\boldsymbol{G} = 8\pi\boldsymbol{T}$ predicts a *non*permanent universe; a dynamic universe; a universe that originated in a "big-bang" explosion, or will be destroyed eventually by contraction to infinite density, or both. Faced with this contradiction between his theory and the firm philosophical beliefs of the day, Einstein weakened; he modified his theory.

The only conceivable modification that does not alter vastly the structure of the theory is to change the lefthand side of the geometrodynamic law $\boldsymbol{G} = 8\pi \boldsymbol{T}$. Recall that the lefthand side is *forced* to be the Einstein tensor, $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta}$, by three assumptions:

(1) $\boldsymbol{G}$ vanishes when spacetime is flat.
(2) $\boldsymbol{G}$ is constructed from the Riemann curvature tensor and the metric and nothing else.
(3) $\boldsymbol{G}$ is distinguished from other tensors that can be built from ***Riemann*** and $\boldsymbol{g}$ by the demands (1) that it be linear in ***Riemann***, as befits any natural measure of curvature; (2) that, like $\boldsymbol{T}$, it be symmetric and of second rank; and (3) that it have an automatically vanishing divergence, $\boldsymbol{\nabla} \cdot \boldsymbol{G} \equiv 0$.

Denote a new, modified lefthand side by "$\boldsymbol{G}$", with quotation marks to avoid confusion with the standard Einstein tensor. To abandon $\boldsymbol{\nabla} \cdot$ "$\boldsymbol{G}$" $\equiv 0$ is impossible on dynamic grounds (see §17.2). To change the symmetry or rank of "$\boldsymbol{G}$" is impossible on mathematical grounds, since "$\boldsymbol{G}$" must be equated to $\boldsymbol{T}$. To let "$\boldsymbol{G}$" be nonlinear in ***Riemann*** would vastly complicate the theory. To construct "$\boldsymbol{G}$" from anything except ***Riemann*** and $\boldsymbol{g}$ would make "$\boldsymbol{G}$" no longer a measure of spacetime geometry and would thus violate the spirit of the theory. After much anguish, one concludes that the assumption which one might drop with least damage to the beauty and spirit of the theory is assumption (1), that "$\boldsymbol{G}$" vanish when spacetime is flat. But even dropping this assumption is painful: (1) although "$\boldsymbol{G}$" might still be in some sense a measure of geometry, it can no longer be a measure of curvature; and (2) flat, empty spacetime will no longer be compatible with the geometrodynamic law ($\boldsymbol{G} \neq 0$ in flat, empty space, where $\boldsymbol{T} = 0$). Nevertheless, these consequences were less painful to Einstein than a dynamic universe.

The only tensor that satisfies conditions (2) and (3) [with (1) abandoned] is the Einstein tensor plus a multiple of the metric:

$$\text{“}G_{\alpha\beta}\text{”} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R + \Lambda g_{\alpha\beta} = G_{\alpha\beta} + \Lambda g_{\alpha\beta}$$

(exercise 17.1; see also exercise 17.4). Thus was Einstein (1917) led to his modified field equation

Einstein's field equation with the cosmological constant

$$\boldsymbol{G} + \Lambda \boldsymbol{g} = 8\pi \boldsymbol{T}. \tag{17.11}$$

The constant Λ he called the "cosmological constant"; it has dimensions cm^{-2}.

Why Einstein abandoned the cosmological constant

The modified field equation, by contrast with the original, admits a static, unchanging universe as one particular solution (see Box 27.5). For this reason, Einstein in 1917 was inclined to place his faith in the modified equation. But thirteen years later Hubble discovered the expansion of the universe. No longer was the cosmological constant necessary. Einstein, calling the cosmological constant "the biggest blunder of my life," abandoned it and returned to his original geometrodynamic law, $\boldsymbol{G} = 8\pi \boldsymbol{T}$ [Einstein (1970)].

A great mistake Λ was indeed!—not least because, had Einstein stuck by his original equation, he could have claimed the expansion of the universe as the most triumphant prediction of his theory of gravity.

A mischievous genie, once let out of a bottle, is not easily reconfined. Many workers in cosmology are unwilling to abandon the cosmological constant. They insist that it be abandoned only after cosmological observations reveal it to be negligibly small. As a modern-day motivation for retaining the cosmological constant, one sometimes rewrites the modified field equation in the form

A modern-day motivation for the cosmological constant: vacuum polarization

$$\boldsymbol{G} = 8\pi[\boldsymbol{T} + \boldsymbol{T}^{(\mathrm{VAC})}], \tag{17.12a}$$

$$\boldsymbol{T}^{(\mathrm{VAC})} \equiv -(\Lambda/8\pi)\boldsymbol{g} \tag{17.12b}$$

and interprets $\boldsymbol{T}^{(\mathrm{VAC})}$ as a stress-energy tensor associated with the vacuum. This viewpoint speculates [Zel'dovich (1967)] that the vacuum polarization of quantum field theory endows the vacuum with the nonzero stress-energy tensor (17.12b), which is completely unobservable except by its gravitational effects. Unfortunately, today's quantum field theory is too primitive to allow a calculation of $\boldsymbol{T}^{(\mathrm{VAC})}$ from first principles. (See, however, exercise 17.5.)

The mass-energy density that the cosmological constant attributes to the vacuum is

$$\rho^{(\mathrm{VAC})} = T^{(\mathrm{VAC})}_{\hat{0}\hat{0}} = +\Lambda/8\pi. \tag{17.13}$$

If $\Lambda \neq 0$, it must at least be so small that $\rho^{(\mathrm{VAC})}$ has negligible gravitational effects $[|\rho^{(\mathrm{VAC})}| < \rho^{(\mathrm{MATTER})}]$ wherever Newton's theory of gravity gives a successful account of observations. The systems of lowest density to which one applies Newtonian theory with some (though not great) success are small clusters of galaxies. Hence, one can place the limit

Observational limit on the cosmological constant

$$|\rho^{(\mathrm{VAC})}| = |\Lambda|/8\pi \lesssim \rho^{(\mathrm{CLUSTER})} \sim 10^{-29}\ \mathrm{g/cm^3} \sim 10^{-57}\ \mathrm{cm^{-2}} \tag{17.14}$$

on the value of the cosmological constant. Evidently, even if $\Lambda \neq 0$, Λ is so small that it is totally unimportant on the scale of a galaxy or a star or a planet or a man or an atom. Consequently it is reasonable to stick with Einstein's original geometrodynamic law ($\boldsymbol{G} = 8\pi\boldsymbol{T}$; $\Lambda = 0$) everywhere, except occasionally when discussing cosmology (Chapters 27–30).

Why one ignores the cosmological constant everywhere except in cosmology

---

EXERCISE

Exercise 17.5. MAGNITUDE OF COSMOLOGICAL CONSTANT

(a) What is the order of magnitude of the influence of the cosmological constant on the celestial mechanics of the solar system if $\Lambda \sim 10^{-57}\ \mathrm{cm^{-2}}$?

(b) Show that the mass-energy density of the vacuum $\rho^{(\mathrm{VAC})} = \Lambda/8\pi \sim 10^{-29}\,\mathrm{g/cm^3}$, corresponding to the maximum possible value of Λ, agrees in very rough magnitude with

$$\frac{\text{rest mass of an elementary particle}}{(\text{Compton wavelength of particle})^3} \times (\text{gravitational fine-structure constant})$$
$$= \frac{m}{(\hbar/m)^3}\frac{m^2}{\hbar} = \frac{m^6}{\hbar^4}$$

[Zel'dovich (1967, 1968)]. This numerology is suggestive, but has not led to any believable derivation of a stress-energy tensor for the vacuum.

§17.4. THE NEWTONIAN LIMIT

Just as quantum mechanics reduces to classical mechanics in the "correspondence limit" of large actions, $I \gg \hbar$, so general relativity reduces to Newtonian theory in the "correspondence limit" of weak gravity and low velocities. (On "correspondence limits," see Box 17.1.) This section elucidates, in some mathematical detail, the correspondence between general relativity and Newtonian theory. It begins with "passive" aspects of gravitation (response of matter to gravity) and then turns to "active" aspects (generation of gravity by matter).

Consider an isolated system—e.g., the solar system—in which Newtonian theory is highly accurate. In order that special relativistic effects not be noticeable, all

Box 17.1 CORRESPONDENCE PRINCIPLES

A. General Remarks and Specific Examples

1. As physics develops and expands, its unity is maintained by a network of correspondence principles, through which simpler theories maintain their vitality by links to more sophisticated but more accurate ones.
   a. Physical optics, with all the new diffraction and interference phenomena for which it accounted, nevertheless also had to account, and did account, for the old, elementary, geometric optics of mirrors and lenses. Geometric optics is recovered from physical optics in the mathematical "correspondence principle limit" in which the wavelength is made indefinitely small in comparison with all other relevant dimensions of the physical system.
   b. Newtonian mechanics is recovered from the mechanics of special relativity in the mathematical "correspondence principle limit" in which all relevant velocities are negligibly small compared to the speed of light.
   c. Thermodynamics is recovered from its successor theory, statistical mechanics, in the mathematical "correspondence principle limit" in which so many particles are taken into account that fluctuations in pressure,

particle number, and other physical quantities are negligible compared to the average values of these parameters of the system.

d. Classical mechanics is recovered from quantum mechanics in the "correspondence principle limit" in which the quantum numbers of the quantum states in question are so large, or the quantities of action that come into play are so great compared to $\hbar$, that wave and diffraction phenomena make negligible changes in the predictions of standard deterministic classical mechanics. Niels Bohr formulated and took advantage of this correspondence principle even before any proper quantum theory existed. He used it to predict approximate values of atomic energy levels and of intensities of spectral lines. He also expounded it as a guide to all physicists, first in searching for a proper version of the quantum theory, and then in elucidating the content of this theory after it was found.

2. In all these examples and others, the newer, more sophisticated theory is "better" than its predecessor because it gives a good description of a more extended domain of physics, or a more accurate description of the same domain, or both.
3. The correspondence between the newer theory and its predecessor (a) gives one the power to recover the older theory from the newer; (b) can be exhibited by straightforward mathematics; and (c), according to the historical record, often guided the development of the newer theory.

B. Correspondence Structure of General Relativity

1. Einstein's theory of gravity has as distinct limiting cases (a) special relativity; (b) the "linearized theory of gravity"; (c) Newton's theory of gravity; and (d) the post-Newtonian theory of gravity. Thus, it has a particularly rich correspondence structure.

a. *Correspondence with special relativity:* General relativity has two distinct kinds of correspondence with special relativity. The first is the limit of vanishing gravitational field everywhere (vanishing curvature); in this limit one can introduce a global inertial frame, set $g_{\mu\nu} = \eta_{\mu\nu}$, and recover completely and precisely the theory of special relativity. The second is local rather than global; it is the demand ("correspondence principle"; "equivalence principle") that in a local inertial frame all the laws of physics take on their special relativistic forms. As was seen in Chapter 16, this puts no restrictions on the metric (except that $g_{\mu\nu} = \eta_{\mu\nu}$ and $g_{\mu\nu,\alpha} = 0$ in local inertial frames); but it places severe constraints on the behavior of matter and fields in the presence of gravity.

b. *Correspondence with Newtonian theory:* In the limit of weak gravitational fields, low velocities, and small pressures, general relativity reduces to Newton's theory of gravity. The correspondence structure is explored mathematically in the text of §17.4.

c. *Correspondence with post-Newtonian theory:* When Newtonian theory is nearly valid, but "first-order relativistic corrections" might be important, one often uses the "post-Newtonian theory of gravity." Chapter 39 expounds the post-Newtonian theory and its correspondence with both general relativity and Newtonian theory.

d. *Correspondence with linearized theory:* In the limit of weak gravitational fields, but possibly large velocities and pressures ($v \sim 1$, $T_{jk} \sim T_{00}$) general relativity reduces to the "linearized theory of gravity". This correspondence is explored in Chapter 18.

Conditions which a system must satisfy for Newton's theory of gravity to be accurate

velocities in the system, relative to its center of mass and also relative to the Newtonian coordinates, must be small compared to the speed of light

$$v \ll 1. \tag{17.15a}$$

As a particle falls from the outer region of the system to the inner region, gravity accelerates it to a kinetic energy $\frac{1}{2}mv^2 \sim |m\Phi|_{\max}$. [Here $\Phi < 0$ is Newton's gravitational potential, so normalized that $\Phi(\infty) = 0$.] The resulting velocity will be small only if

$$|\Phi| \ll 1. \tag{17.15b}$$

Internal stresses in the system also produce motion—e.g., sound waves. Such waves have characteristic velocities of the order of $|T^{ij}/T^{00}|^{1/2}$—for example, the speed of sound in a perfect fluid is

$$v = (dp/d\rho)^{1/2} \sim (p/\rho)^{1/2} \sim |T^{ij}/T^{00}|^{1/2}.$$

In order that these velocities be small compared to the speed of light, all stresses must be small compared to the density of mass-energy

$$|T^{ij}|/T^{00} = |T^{ij}|/\rho \ll 1. \tag{17.15c}$$

When, and only when conditions (17.15) hold, one can expect Newtonian theory to describe accurately the system being studied. Correspondence of general relativity with Newtonian theory for gravity in a passive role then demands that the geodesic world lines of freely falling particles reduce to the Newtonian world lines

$$d^2x^i/dt^2 = -\partial\Phi/\partial x^i. \tag{17.16}$$

Moreover, they must reduce to this form in *any* relativistic coordinate system where the source and test particles have low velocities $v \ll 1$, and where coordinate lengths and times agree very nearly with the lengths and times of the Newtonian coordinates—which in turn are proper lengths and times as measured by rods and clocks. Thus, the relevant coordinates (called "*Galilean*" *or* "*Newtonian*" *coordinates*) are ones in which

"Newtonian coordinates" defined

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1, \qquad |v^j| = |dx^j/dt| \ll 1 \tag{17.17}$$

(weak gravitational field; nearly inertial coordinates; low velocities). In such a coordinate system, the geodesic world lines of test particles have the form

$$\begin{aligned}
\frac{d^2x^i}{dt^2} &= \frac{d^2x^i}{d\tau^2} && (\text{since } dt/d\tau \approx 1 \text{ when } |h_{\mu\nu}| \ll 1 \text{ and } |v^j| \ll 1)\\
&= -\Gamma^i{}_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} && (\text{geodesic equation})\\
&= -\Gamma^i{}_{00} && (\text{since } dt/d\tau \approx 1 \text{ and } |dx^j/d\tau| \ll 1)\\
&= -\Gamma_{i00} && (\text{since } g_{\mu\nu} \approx \eta_{\mu\nu})\\
&= \frac{1}{2}h_{00,i} - h_{0i,0} && (\text{equation for } \Gamma_{\alpha\beta\gamma} \text{ in terms of } g_{\alpha\beta,\gamma})\\
&= \frac{1}{2}h_{00,i} && \left(\begin{array}{l}\text{all velocities small compared to } c \text{ implies time}\\ \text{derivatives small compared to space derivatives}\\ \text{—i.e., } h_{\alpha\beta,0} \sim vh_{\alpha\beta,i}\end{array}\right).
\end{aligned}$$

These geodesic world lines do, indeed, reduce to those of Newtonian theory [equation (17.16)] if one makes the identification

$$\Gamma^i{}_{00} = -\frac{1}{2}h_{00,i} = \Phi_{,i}. \tag{17.18}$$

Together with the boundary conditions $\Phi(r = \infty) = 0$ and $h_{\mu\nu}(r = \infty) = 0$ (coordinates Lorentz far from the source), this identification implies $h_{00} = -2\Phi$; i.e.,

Einstein gravity reduces to Newton gravity only if, in Newtonian coordinates, $g_{00} = -1 - 2\Phi$

$$g_{00} = -1 - 2\Phi \text{ for nearly Newtonian systems in Newtonian coordinates.} \tag{17.19}$$

Note that the correspondence tells one the form of h_{00} for nearly Newtonian systems, but not the forms of the other components of the metric perturbation. In fact, the other $h_{\mu\nu}$ could perfectly well be of the same order of magnitude as $h_{00} \sim \Phi$, without influencing the world lines of slowly moving particles, because they always enter the geodesic equation multiplied by the small numbers v or v^2, or differentiated by t rather than by x^i. The forms of the other $h_{\mu\nu}$ and their small corrections to the Newtonian motion will be explored in Chapters 18, 39, and 40.

The correspondence between Einstein theory and Newton theory for all "passive" aspects of gravity

The relation $g_{00} = -1 - 2\Phi$ is the mathematical embodiment of the correspondence between general relativity theory and Newtonian theory for passive aspects of gravity. Together with the "validity conditions" (17.15, 17.17), it is a foundation from which one can derive all other aspects of the correspondence for "passive gravity," including the relation

$$R^i{}_{0j0} = \partial^2\Phi/\partial x^i\,\partial x^j \tag{17.20}$$

(exercise 17.6). Alternatively, all other aspects of this correspondence can be derived by direct comparison of Newton's predictions with Einstein's. For example, to derive equation (17.20), examine the relative acceleration of two test particles, one at $x^i + \xi^i$ and the other at x^i. According to Newton

$$\begin{aligned}\frac{d^2\xi^i}{dt^2} &= \frac{d^2(x^i + \xi^i)}{dt^2} - \frac{d^2x^i}{dt^2} \\ &= -\left.\frac{\partial\Phi}{\partial x^i}\right|_{\text{at } x^j + \xi^j} + \left.\frac{\partial\Phi}{\partial x^i}\right|_{\text{at } x^j} = \frac{-\partial^2\Phi}{\partial x^i\,\partial x^j}\xi^j.\end{aligned}$$

For comparison, Einstein predicts (equation of geodesic deviation)

$$\frac{D^2\xi^i}{d\tau^2} \underset{\uparrow}{=} \frac{d^2\xi^i}{dt^2} = -R^i{}_{0j0}\xi^j.$$

[by conditions (17.15) and (17.17)]

Direct comparison gives relation (17.20).

Turn now from correspondence for passive aspects of gravity to correspondence for active aspects. According to Einstein, mass generates gravity (spacetime curvature) by the geometrodynamic law $\boldsymbol{G} = 8\pi\boldsymbol{T}$. Apply this law to a nearly Newtonian system, and by the chain of reasoning that preceeds equation (17.8) derive the relation

The Newtonian limit of the Einstein field equation is $\nabla^2\Phi = 4\pi\rho$

$$R_{00} = 4\pi\rho. \tag{17.21}$$

Combine with the contraction of (17.20),

$$R_{00} = R^i{}_{0i0} + \underset{\substack{\uparrow \\ 0}}{R^0{}_{000}} = \partial^2\Phi/\partial x^i\,\partial x^i = \nabla^2\Phi,$$

and thereby obtain Newton's equation for the generation of gravity by mass

$$\nabla^2\Phi = 4\pi\rho. \tag{17.22}$$

Thus, Einstein's field equation reduces to Newton's field equation in the Newtonian limit.

The correspondence between Newton and Einstein, although clear and straightforward as outlined above, is even more clear and straightforward when Newton's theory of gravity is rewritten in Einstein's language of curved spacetime (Chapter 12; exercise 17.7).

EXERCISES

Exercise 17.6. RAMIFICATIONS OF CORRESPONDENCE FOR GRAVITY IN A PASSIVE ROLE

From the correspondence relation $g_{00} = -1 - 2\Phi$, and from conditions (17.15) and (17.17) for Newtonian physics, derive the correspondence relations

$$\Gamma^i{}_{00} = \partial\Phi/\partial x^i, \qquad R^i{}_{0j0} = \partial^2\Phi/\partial x^i\,\partial x^j.$$

Exercise 17.7. CORRESPONDENCE IN THE LANGUAGE OF CURVED SPACETIME [Track 2]

Exhibit the correspondence between the Einstein theory and Cartan's curved-spacetime formulation of Newtonian theory (Chapter 12).

---

§17.5. AXIOMATIZE EINSTEIN'S THEORY?

Find the most compact and reasonable axiomatic structure one can for general relativity? Then from the axioms derive Einstein's field equation,

$$\boldsymbol{G} = 8\pi\boldsymbol{T}?$$

That approach would follow tradition. However, it may be out of date today. More than half a century has gone by since November 25, 1915. For all that time the equation has stood unchanged, if one ignores Einstein's temporary "aberration" of adding the cosmological constant. In contrast the derivations have evolved and become more numerous and more varied. In the beginning axioms told what equation is acceptable. By now the equation tells what axioms are acceptable. Box 17.2 sketches a variety of sets of axioms, and the resulting derivations of Einstein's equation.

There are many ways (Box 17.2) to derive the Einstein field equation

(*continued on page 429*)

Box 17.2 SIX ROUTES TO EINSTEIN'S GEOMETRODYNAMIC LAW OF THE EQUALITY OF CURVATURE AND ENERGY DENSITY ("EINSTEIN'S FIELD EQUATION")

[Recommended to the attention of Track-1 readers are only route 1 (automatic conservation of the source, plus correspondence with Newtonian theory) and route 2 (Hilbert's variational principle); and even Track-2 readers are advised to finish the rest of this chapter before they study route 3 (physics on a spacelike slice), route 4 (going from superspace to Einstein's equation), route 5 (field of spin 2 in an "unobservable flat spacetime" background), and route 6 (gravitation as an elasticity of space that arises from particle physics).]

1. Model geometrodynamics after electrodynamics and treat "automatic conservation of the source" and correspondence with the Newtonian theory of gravity as the central considerations.
   a. Particle responds in electrodynamics to field; in general relativity, to geometry.
   b. The potential for the electromagnetic field is the 4-vector $\boldsymbol{A}$ (components A_μ). The potential for the geometry is the metric tensor $\boldsymbol{g}$ (components $g_{\mu\nu}$).
   c. The electromagnetic potential satisfies a wave equation with source term (4-current) on the right,

$$\left(\frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}\right)^{;\nu} = 4\pi j_\mu, \tag{1}$$

   so constructed that conservation of the source, $j_\mu{}^{;\mu} = 0$, is automatic (consequence of an identity fulfilled by the lefthand side). By analogy, the geometrodynamic potential must also satisfy a wave equation with source term (stress-energy tensor) on the right,

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{2}$$

   so constructed that conservation of the source, $T_{\mu\nu}{}^{;\nu} = 0$ (Chapter 16) is "automatic." This conservation is automatic here because the lefthand side of the equation is a tensor (the Einstein tensor; see Box 8.6 or Chapter 15), built from the metric components and their second derivatives, that fulfills the identity $G_{\mu\nu}{}^{;\nu} \equiv 0$.
   d. No other tensor which (1) is linear in the second derivatives of the metric components, (2) is free of higher derivatives, and (3) vanishes in flat spacetime, satisfies such an identity.
   e. The constant of proportionality (8π) is fixed by the choice of units [here geometric; see Box 1.8] and by the requirement ("correspondence with Newtonian theory") that a test particle shall oscillate back and forth through a collection of matter of density ρ, or revolve in circular orbit around that collection of matter, at a circular frequency given by $\omega^2 = (4\pi/3)\rho$ (Figure

Box 17.2 (*continued*)

1.12). The foregoing oversimplifies, and omits Einstein's temporary false turns, but otherwise summarizes the reasoning he pursued in arriving at his field equation. This reasoning is spelled out in more detail in the text of Chapter 17.

2. Take variational principle as central.
   a. Construct out of the metric components the only scalar that exists that (1) is linear in the second derivatives of the metric tensor, (2) contains no higher derivatives, and (3) vanishes in flat spacetime: namely, the Riemann scalar curvature invariant, R.
   b. Construct the invariant integral,
   $$I = \frac{1}{16\pi}\int_{\Omega} R(-g)^{1/2}\, d^4x. \tag{3}$$
   c. Make small variations, $\delta g^{\mu\nu}$, in the metric coefficients $g^{\mu\nu}$ in the interior of the four-dimensional region Ω, and find that this integral changes by the amount
   $$\delta I = \frac{1}{16\pi}\int_{\Omega} G_{\mu\nu}\, \delta g^{\mu\nu}(-g)^{1/2}\, d^4x. \tag{4}$$
   d. Demand that I should be an extremum with respect to the choice of geometry in the region interior to Ω ($\delta I = 0$ for arbitrary $\delta g^{\mu\nu}$; "principle of extremal action").
   e. Thus arrive at the Einstein field equation for empty space,
   $$G_{\mu\nu} = 0. \tag{5}$$
   f. The continuation of the reasoning leads to the identity
   $$G_{\mu\nu}{}^{;\nu} = 0.$$
   Chapter 21, on the variational principle, gives more detail and takes up the additional term that appears on the righthand side of (5) when matter or fields or both are present.
   g. This approach goes back to David Hilbert (1915). No route to the field equations is quicker. Moreover, it connects immediately (see the following section here, 2′) with the quantum principle of the "democracy of all histories" [Feynman (1942); Feynman and Hibbs (1965)]. The variational principle is spelled out in more detail in Chapter 21.

2′. An aside on the meaning of the classical action integral for the real world of quantum physics.
   a. A "history of geometry," H, is a spacetime, that is to say, a four-dimensional manifold with four-dimensional $-+++$ Riemann metric that (1) reduces on one spacelike hypersurface ("hypersurface of simultaneity") to a specified "initial value 3-geometry," A, with positive definite metric and (2) reduces on

another spacelike hypersurface to a specified "final value 3-geometry," B, also with positive definite metric.

b. The classical variational principle of Hilbert, as reformulated by Arnowitt, Deser, and Misner, provides a prescription for the dynamical path length, I_H, of any conceivable history H, classically allowed or not, that connects A and B (see Chapter 21 for a fuller statement for what can and must be specified on the initial hypersurface of simultaneity, and on the final one, and for alternative choices of the integrand in the action principle).

c. Classical physics says that a history H is allowed only if it extremizes the dynamic path length I as compared to all nearby histories. Quantum physics says that all histories occur with equal probability amplitude, in the following sense. The probability amplitude for "the dynamic geometry of space to transit from A to B" by way of the history H with action integral I_H, and by way of histories that lie within a specified infinitesimal range, $\mathscr{D}H$, of the history H, is given by the expression

$$\begin{pmatrix} \text{probability amplitude} \\ \text{to transit from } A \text{ to} \\ B \text{ by way of history } H \\ \text{and histories lying} \\ \text{within the range } \mathscr{D}H \\ \text{about } H \end{pmatrix} \sim \exp\,(iI_H/\hbar)N\mathscr{D}H. \tag{6}$$

Here the normalization factor, N, is the same for all conceivable histories H, allowed or not, that lead from A to B ("principle of democracy of histories"). The quantum of angular momentum, $\hbar = h/2\pi$, expressed in geometric units, has the value

$$\hbar = \hbar_{\text{conv}} G/c^3 = (L^*)^2, \tag{7}$$

where L^* is the Planck length, $L^* = 1.6 \times 10^{-33}$ cm.

d. The classically allowed history receives "preference without preference." That history, and histories H that differ from it so little that $\delta I = I_H - I_{class}$ is only of the order $\hbar$ and less, give contributions to the probability amplitude that interfere constructively. In contrast, destructive interference effectively wipes out the contribution (to the probability amplitude for a transition) that comes from histories that differ more from the classically allowed history. Thus there are quantum fluctuations in the geometry, but they are fluctuations of limited magnitude. The smallness of $\hbar$ ensures that the scale of these fluctuations is unnoticeable at everyday distances (see the further discussion in Chapters 43 and 44). In this sense classical geometrodynamics is a good approximation to the geometrodynamics of the real world of quantum physics.

3. "Physics on a spacelike slice or hypersurface of simultaneity," again with electromagnetism as the model.

a. Say over and over "lines of magnetic force never end" and come out with half of Maxwell's equations. Say over and over "lines of electric force end

The rest of this chapter is Track 2. No previous track-2 material is needed as preparation for it, nor is it necessary preparation for any later chapter, but it will be helpful in Chapter 21 (initial-value equations and variational principle) and in Chapter 39 (other theories of gravity).

Box 17.2 (*continued*)

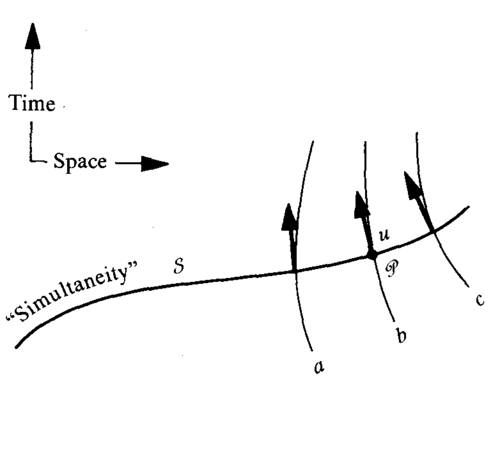
only on charge" and arrive at the other half of Maxwell's equations. Similarly, say over and over

$$\begin{pmatrix}\text{intrinsic}\\ \text{curvature}\\ \text{scalar}\end{pmatrix} + \begin{pmatrix}\text{extrinsic}\\ \text{curvature}\\ \text{scalar}\end{pmatrix} = 16\pi \begin{pmatrix}\text{local density}\\ \text{of mass-}\\ \text{energy}\end{pmatrix} \tag{8}$$

and end up with all ten components of Einstein's equation. To "say over and over" is an abbreviation for demanding that the stated principles hold on every spacelike slice through every event of spacetime.

b. Spell out explicitly this "spacelike-slice formulation" of the equations of Maxwell and Einstein. Consider an arbitrary point of spacetime, $\mathcal{P}$ ("event"), and an arbitrary "simultaneity" $\mathcal{S}$ through $\mathcal{P}$ (hypersurface of simultaneity; spacelike slice through spacetime). Magnetic lines of force run about throughout $\mathcal{S}$, but nowhere is even a single one of them permitted to end. Recall (§3.4) that the demand "lines of magnetic force never end," when imposed on *all* reference frames at $\mathcal{P}$ (for all choices of the "simultaneity" $\mathcal{S}$), guarantees not only $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$, but also $\boldsymbol{\nabla} \times \boldsymbol{E} + \partial \boldsymbol{B}/\partial t = 0$. Similarly (§3.4) the demand that "electric lines of force never end except on electric charge," $\boldsymbol{\nabla} \cdot \boldsymbol{E} = 4\pi J^0$, when imposed on all "simultaneities" through $\mathcal{P}$, guarantees the remaining Maxwell equation $\boldsymbol{\nabla} \times \boldsymbol{B} = \partial \boldsymbol{E}/\partial t + 4\pi \boldsymbol{J}$.

c. Each simultaneity $\mathcal{S}$ through $\mathcal{P}$ has its own slope and curvature. The possibility of different slopes (different local Lorentz frames at $\mathcal{P}$) is essential for deriving all of Maxwell's equations from the requirements of conservation of flux. Relevant though the slope thus is, the curvature of the hypersurface $\mathcal{S}$ never matters for the analysis of electromagnetism. It does matter, however, for any analysis of gravitation modeled on the foregoing treatment of electromagnetism.



"Simultaneity" $\mathcal{S}$ (spacelike hypersurface or "slice through spacetime") that cuts through event $\mathcal{P}$. The "simultaneity" may be considered to be defined by a set of "observers" a,b,c, ... Their world lines cross the simultaneity orthogonally, and their clocks all read the same proper time at the instant of crossing. Another simultaneity through $\mathcal{P}$ may have at $\mathcal{P}$ a different curvature or a different slope or both; and it is defined by a different band of observers, with other wrist watches.

d. "Mass-energy curves space" is the central principle of gravitation. To spell out this principle requires one to examine in succession the terms "space" and "curvature of space" and "density of mass-energy in a given region of space." "Space" means spacelike hypersurface; or, more specifically, a hypersurface of simultaneity $\mathcal{S}$ that includes the point $\mathcal{P}$ where the physics is under examination.

e. Denote by $\boldsymbol{u}$ the 4-vector normal to $\mathcal{S}$ at $\mathcal{P}$. Then the density of mass-energy in the spacelike hypersurface $\mathcal{S}$ at $\mathcal{P}$ is

$$\rho = u^{\alpha} T_{\alpha\beta} u^{\beta}, \tag{9}$$

in accordance with the definition of the stress-energy tensor given in Chapter 5.

f. This density is a single number, dependent on the inclination of the slice one cuts through spacetime, but independent of how curved one cuts this slice. If it is to be equated to "curvature of space," that curvature must also be independent of how curved one cuts the slice.

g. Conclude that the geometric quantity, "curvature of space," must (1) be a single number (a scalar) that (2) depends on the inclination $\boldsymbol{u}$ of the cut one makes through spacetime at $\mathcal{P}$ in constructing the hypersurface $\mathcal{S}$, but (3) must be unaffected by how one curves his cut. The demand made here appears paradoxical. One seems to be asking for a measure of curvature that is independent of curvature!

h. A closer look discloses that three distinct ideas come into consideration here. One is the scalar curvature invariant ${}^{(3)}R$ of the 3-geometry intrinsic to the hypersurface $\mathcal{S}$ at $\mathcal{P}$: "intrinsic" in the sense that it is defined by, and depends exclusively on, measurements of distance made within the hypersurface. The second is the "extrinsic curvature" of this 3-geometry relative to the 4-geometry of the enveloping spacetime ("how curved one cuts his slice"; see Box 14.1 for more on the distinction between extrinsic and intrinsic curvature). The third is the curvature of the four-dimensional spacetime itself, "normal to $\boldsymbol{u}$," in some sense yet to be more closely defined. This is the quantity that is independent of how curved one cuts his slice. It is the quantity that is to be identified, up to a factor that depends on the choice of units, with the density of mass-energy.

2

Box 17.2 (*continued*)

i. These three quantities are related in the following way:

(scalar curvature invariant, ${}^{(3)}R$, of the 3-geometry intrinsic to the spacelike hypersurface $\mathcal{S}$, a quantity dependent on "how curved one cuts the slice")

\+ (a correction term that (a) depends only on the "extrinsic curvature" $K_{\alpha\beta}$ (Box 14.1 and Chapter 21) of the hypersurface relative to the four-dimensional geometry in which it is imbedded, and (b) is so calculated (a uniquely determinate calculation) that the sum of this correction term and ${}^{(3)}R$ is independent of "how curved one cuts his slice," and (c) has the precise value $(\mathrm{Tr}\,\boldsymbol{K})^2 - \mathrm{Tr}\,\boldsymbol{K}^2 \equiv (K_\alpha{}^\alpha)^2 - K_{\alpha\beta}K^{\alpha\beta}$)

= (a measure of the curvature of spacetime that depends on the 4-geometry of the spacetime and on the inclination $\boldsymbol{u}$ of the spacelike slice $\mathcal{S}$ cut through spacetime, but is independent, by construction, of "how curved one cuts the slice")

= (a scalar quantity that (a) is completely defined by what has just been said and (b) can therefore be calculated in all completeness by standard differential geometry (details in Chapter 21))

= ($2u^\alpha G_{\alpha\beta} u^\beta$, where $G_{\alpha\beta}$ is the Einstein curvature tensor of equation 8.49 and Box 8.6)

= 2 (a quantity interpreted in Track 2, Chapter 15, as the "moment of rotation" associated with a unit element of 3-volume located at $\mathcal{P}$ in the hypersurface orthogonal to $\boldsymbol{u}$) (10)

j. Conclude that the central principle, "mass-energy curves space," translates to the formula

$$ {}^{(3)}R + (\mathrm{Tr}\,\boldsymbol{K})^2 - \mathrm{Tr}\,\boldsymbol{K}^2 = 16\pi\rho, \tag{11}$$

or, in shorthand form,

$$\begin{pmatrix}\text{moment of}\\ \text{rotation}\end{pmatrix} = \begin{pmatrix}\text{intrinsic}\\ \text{curvature}\end{pmatrix} + \begin{pmatrix}\text{extrinsic}\\ \text{curvature}\end{pmatrix} = \begin{pmatrix}\text{density of}\\ \text{mass-energy}\end{pmatrix}, \tag{12}$$

valid for every spacelike slice through spacetime at any arbitrary point $\mathcal{P}$.

k. All of Einstein's geometrodynamics is contained in this statement as truly as all of Maxwell's electrodynamics is contained in the statement that the number of lines of force that end in an element of volume is equal to 4π times the amount of charge in that element of volume. The factor 16π is appropriate for the geometric system of units in use in this book (density ρ in cm^{-2} given by $G/c^2 = 0.742 \times 10^{-28}$ cm/g multiplied by the density ρ_{conv} expressed in the conventional units of g/cm^3).

l. Reexpress the principle that "mass-energy curves space" in the form

$$2u^\alpha G_{\alpha\beta}u^\beta = 16\pi u^\alpha T_{\alpha\beta}u^\beta. \tag{13}$$

Demand that this equation should hold for every simultaneity that cuts through $\mathcal{P}$, whatever its "inclination" $\boldsymbol{u}$.

m. Conclude that the coefficients on the two sides of (13) must agree; thus,

$$G_{\alpha\beta} = 8\pi T_{\alpha\beta}, \tag{14}$$

Einstein's equation in the language of components; or, in the language of abstract geometric quantities,

$$\boldsymbol{G} = 8\pi \boldsymbol{T}. \tag{15}$$

4. Going from superspace to Einstein's equation rather than from Einstein's equation to superspace.
   a. A fourth route to Einstein's equation starts with the advanced view of geometrodynamics that is spelled out in Chapter 43. One notes there that the dynamics of geometry unfolds in superspace. Superspace has an infinite number of dimensions. Any one point in superspace describes a complete 3-geometry, ${}^{(3)}\mathcal{G}$, with all its bumps and curvatures. The dynamics of geometry leads from point to point in superspace.
   b. Like the dynamics of a particle, the dynamics of geometry lends itself to distinct but equivalent mathematical formulations, associated with the names of Lagrange, of Hamilton, and of Hamilton and Jacobi. Of these the most convenient for the present analysis is the last ("H-J").
   c. In the problem of one particle moving in one dimension under the influence of a potential $V(x)$, the H-J equation reads

$$\underbrace{-\frac{\partial S}{\partial t}}_{\text{total energy}} = \underbrace{\frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2}_{\text{kinetic energy}} + V(x). \tag{16}$$

Box 17.2 (*continued*)

It has the solution

$$S_E(x,t) = -Et + \int^x [2m(E - V)]^{1/2}\, dx. \tag{17}$$

Out of this solution one reads the motion by applying the "condition of constructive interference,"

$$\frac{\partial S_E(x,t)}{\partial E} = 0 \tag{18}$$

(one equation connecting the two quantities x and t; for more on the condition of constructive interference and the H-J method in general, see Boxes 25.3 and 25.4).

d. In the corresponding equation for the dynamics of geometry, one deals with a function $S = S({}^{(3)}\mathcal{G})$ of the 3-geometry. It depends on the 3-geometry itself, and not on the vagaries of one's choice of coordinates or on the corresponding vagaries in the metric coefficients of the 3-geometry,

$$ds^2 = {}^{(3)}g_{mn}\, dx^m\, dx^n \tag{19}$$

(${}^{(3)}$ to indicate 3-geometry omitted hereafter for simplicity). This function obeys the H-J equation [the analog of (16)]

$$-(16\pi)^2 \frac{1}{2g}(g_{im}g_{jn} + g_{in}g_{jm} - g_{ij}g_{mn})\frac{\delta S}{\delta g_{ij}}\frac{\delta S}{\delta g_{mn}} + {}^{(3)}R = 16\pi\rho. \tag{20}$$

e. Out of this equation for the dynamics of geometry in superspace one can deduce the Einstein field equation by reasoning similar to that employed in going from (17) to (18) (Gerlach 1969).
f. It would appear that one must break new ground, and establish new foundations, if one is to find out how to regard the "Einstein-Hamilton-Jacobi equation" (20) as more basic than the Einstein field equation that one derives from it. [Since done, by Hojman, Kuchař, and Teitelboim (1973 preprint).]

5. Einstein's geometrodynamics viewed as the standard field theory for a field of spin 2 in an "unobservable flat spacetime" background.
   a. This approach to Einstein's field equation has a long history, references to which will be found in §7.1 and §18.1. (Further discussion of this approach will be found in those two sections and in Box 7.1, exercise 7.3, and Box 18.1).
   b. The following summary is quoted from Deser (1970): "We wish to give a simple physical derivation of the nonlinearity . . . , using a now familiar argument . . . leading from the linear, massless, spin-2 field to the full Einstein equations . . . .

c. "The Einstein equations may be derived nongeometrically by noting that the free, massless, spin-2 field equations,

$$R^L{}_{\mu\nu}(\phi) - \frac{1}{2} R^L{}_{\alpha\alpha}(\phi)\eta_{\mu\nu} \equiv G^L{}_{\mu\nu}(\phi) \equiv [(\eta_{\mu\alpha}\eta_{\nu\beta} - \eta_{\mu\nu}\eta_{\alpha\beta})\Box$$

$$+ \eta_{\mu\nu}\partial_\alpha\partial_\beta + \eta_{\alpha\beta}\partial_\mu\partial_\nu - \eta_{\mu\alpha}\partial_\nu\partial_\beta - \eta_{\nu\beta}\partial_\mu\partial_\alpha]\phi_{\alpha\beta} = 0, \quad (21)$$

whose source is the matter stress-tensor $T_{\mu\nu}$, must actually be coupled to the *total* stress-tensor, including that of the ϕ-field itself. That is, while the free-field equations (21) are of course quite consistent as they stand, [they are not] when there is a dynamic system's $T_{\mu\nu}$ as a source. For then the left side, which is identically divergenceless, is inconsistent with the right, since the coupling implies that $T^{\mu\nu}{}_{,\nu}$, as computed from the matter equations of motion, is no longer conserved.

d. "To remedy this [violation of the principle of conservation of momentum and energy], the stress tensor ${}^{(2)}\theta_{\mu\nu}$ arising from the quadratic Lagrangian ${}^{(2)}L$ responsible for equation (21) is then inserted on the right.

e. "But the Lagrangian ${}^{(3)}L$ leading to these modified equations is then cubic, and itself contributes a cubic ${}^{(3)}\theta_{\mu\nu}$.

f. "This series continues indefinitely, and sums (if properly derived!) to the full nonlinear Einstein equations, $G_{\mu\nu}$ ([calculated from] $\eta_{\alpha\beta} + \phi_{\alpha\beta}) = -\kappa T_{\mu\nu}$ [$+ 8\pi T_{\mu\nu}$ in the geometric units and sign conventions of this book], which are an infinite series in the deviation $\phi_{\mu\nu}$ of the metric $g_{\mu\nu}$ from its Minkowskian value $\eta_{\mu\nu}$.

g. Once the iteration is begun (whether or not a $T_{\mu\nu}$ is actually present), it must be continued to all orders, since conservation only holds for the full series

$\sum_{n=2}^{\infty} {}^{(n)}\theta_{\mu\nu}$. Thus, the theory is either left in its (physically irrelevant) free linear

form (21), or it *must* be an infinite series."

h. For details, see Deser (1970); the paper goes on (1) to take advantage of a well-chosen formalism (2) to rearrange the calculation, and thus (3) to "derive the full Einstein equations, on the basis of the same self-coupling requirement, but with the advantages that the full theory emerges in closed form with just one added (cubic) term, rather than as an infinite series."

i. Deser summarizes the analysis at the end thus: "Consistency has therefore led us to universal coupling, which implies the equivalence principle. It is at this point that the geometric interpretation of general relativity arises, since *all matter* now moves in an effective Riemann space of metric $\mathfrak{g}^{\mu\nu} \equiv \eta^{\mu\nu} + h^{\mu\nu}$. . . . [The] initial flat 'background' space is no longer observable." In other words, this approach to Einstein's field equation can be summarized as "curvature without curvature" or—equally well—as "flat spacetime without flat spacetime"!

Box 17.2 (*continued*)

6. Sakharov's view of gravitation as an elasticity of space that arises from particle physics.
   a. The resistance of a homogeneous isotropic solid to deformation is described by two elastic constants, Young's modulus and Poisson's ratio.

   b. The resistance of space to deformation is described by one elastic constant, the Newtonian constant of gravity. It makes its appearance in the action principle of Hilbert

$$
\begin{aligned}
I = \frac{1}{16\pi G}\int {}^{(4)}R(-g)^{1/2}\,d^4x \\
+ \int (L_{\text{matter}} + L_{\text{fields}})(-g)^{1/2}\,d^4x = \text{extremum}.
\end{aligned}
\tag{22}
$$

   c. According to the historical records, it was first learned how many elastic constants it takes to describe a solid from microscopic molecular models of matter (Newton, Laplace, Navier, Cauchy, Poisson, Voigt, Kelvin, Born), not from macroscopic considerations of symmetry and invariance. Thus, count the energy stored up in molecular bonds that are deformed from natural length or natural angle or both. Arrive at an expression for the energy of deformation per unit volume of the elastic material of the form

$$e = A(\mathrm{Tr}\,\boldsymbol{s})^2 + B\,\mathrm{Tr}(\boldsymbol{s}^2). \tag{23}$$

   Here the strain tensor

$$s_{mn} = \frac{1}{2}\left(\frac{\partial \xi_m}{\partial x^n} + \frac{\partial \xi_n}{\partial x^m}\right) \tag{24}$$

   measures the strain produced in the elastic medium by motion of the typical point that was at the location x^m to the location $x^m + \xi^m(x)$. The constants A and B are derived out of microscopic physics. They fix the values of the two elastic constants of the macroscopic theory of elasticity.
   d. Andrei Sakharov (1967) (*the* Andrei Sakharov) has proposed a similar microscopic foundation for gravitation or, as he calls it, the "metric elasticity of space." He identifies the action term of Einstein's geometrodynamics [the first term in (22)] "with the change in the action of quantum fluctuations of the vacuum [associated with the physics of particles and fields and brought about] when space is curved."
   e. Sakharov notes that present-day quantum field theory "gets rid by a renormalization process" of an energy density in the vacuum that would formally be infinite if not removed by this renormalization. Thus, in the standard analysis of the degrees of freedom of the electromagnetic field in flat space, one counts the number of modes of vibration per unit volume in the range

of circular wave numbers from k to $k + dk$ as $(2 \cdot 4\pi/8\pi^3)k^2\, dk$. Each mode of oscillation, even at the absolute zero of temperature, has an absolute irreducible minimum of "zero-point energy of oscillation," $\frac{1}{2}h\nu = \frac{1}{2}\hbar ck$ [the fluctuating electric field associated with which is among the most firmly established of all physical effects. It acts on the electron in the hydrogen atom in supplement to the electric field caused by the proton alone, and thereby produces most of the famous Lamb-Retherford shift in the energy levels of the hydrogen atom, as made especially clear by Welton (1948) and Dyson (1954)]. The totalized density of zero-point energy of the electromagnetic field per unit volume of spacetime (units: cm^4) formally diverges as

$$(\hbar/2\pi^2)\int_0^\infty k^3\, dk. \tag{25}$$

Equally formally this divergence is "removed" by "renormalization" [for more on renormalization see, for example, Hepp (1969)].

f. Similar divergences appear when one counts up formally the energy associated with other fields and with vacuum fluctuations in number of pairs of electrons, μ-mesons, and other particles in the limit of quantum energies large in comparison with the rest mass of any of these particles. Again these divergences in formal calculations are "removed by renormalization."

g. Removed by renormalization is a contribution not only to the energy density, and therefore to the stress-energy tensor, but also to the total Lagrange function $\mathcal{L}$ of the variational principle for all these fields and particles,

$$I = \int \mathcal{L}\, d^4x = \text{extremum}. \tag{26}$$

h. Curving spacetime alters all these energies, Sakharov points out, extending an argument of Zel'dovich (1967). Therefore the process of "renormalization" or "subtraction" no longer gives zero. Instead, the contribution of zero-point energies to the Lagrangian, expanded as a power series in powers of the curvature, with numerical coefficients $A, B, \ldots$ of the order of magnitude of unity, takes a form simplified by Ruzmaikina and Ruzmaikin (1969) to the following:

$$\begin{aligned}\mathcal{L}(R) = A\hbar \int k^3\, dk &+ B\hbar^{(4)}R \int k\, dk \\ &+ \hbar[C({}^{(4)}R)^2 + DR^{\alpha\beta}R_{\alpha\beta}]\int k^{-1}\, dk \\ &+ \text{(higher-order terms)}.\end{aligned} \tag{27}$$

[For the alteration in the number of standing waves per unit frequency in a curved manifold, see also Berger (1966), Sakharov (1967), Hill in De Witt (1967c), Polievktov-Nikoladze (1969), and Berger, Gauduchon, and Mazet (1971).]

i. Renormalization physics argues that the first term in (27) is to be dropped. The second term, Sakharov notes, is identical in form to the Hilbert action

Box 17.2 (*continued*)

principle, equation (3) above, with the exception that there the constant that multiplies the Riemann scalar curvature invariant is $-c^3/16\pi G$ (in conventional units), whereas here it is $B\hbar\int k\,dk$ (in the same conventional units). The higher order terms in (27) lead to what Sakharov calls "corrections . . . to Einstein's equations."

j. Overlooking these corrections, one evidently obtains the action principle of Einstein's theory when one insists on the equality

$$G = \begin{pmatrix} \text{Newtonian} \\ \text{constant of gravity} \end{pmatrix} = \frac{c^3}{16\pi B\hbar\int k\,dk}. \tag{28}$$

With B a dimensionless numerical factor of the order of unity, it follows, Sakharov argues, that the effective upper limit or "cutoff" in the formally divergent integral in (28) is to be taken to be of the order of magnitude of the reciprocal Planck length [see equation (7)],

$$k_{\text{cut off}} \sim (c^3/\hbar G)^{1/2} = 1/L^* = 1/1.6 \times 10^{-33}\ \text{cm}. \tag{29}$$

In effect Sakharov is saying (1) that field physics suffers a sea change into something new and strange for wavelengths less than the Planck length, and for quantum energies of the order of $\hbar c k_{\text{cutoff}} \sim 10^{28}$ eV or 10^{-5} g or more; (2) that in consequence the integral $\int k\,dk$ is cut off; and (3) that the value of this cutoff, arising purely out of the physics of fields and particles, governs the value of the Newtonian constant of gravity, G.

k. In this sense, Sakharov's analysis suggests that gravitation is to particle physics as elasticity is to chemical physics: merely a statistical measure of residual energies. In the one case, molecular bindings depend on departures of molecule-molecule bond lengths from standard values. In the other case, particle energies are affected by curvatures of the geometry.

l. Elasticity, which looks simple, gets its explanation from molecular bindings, which are complicated; but molecular bindings, which are complicated, receive their explanation in terms of Schrödinger's wave equation and Coulomb's law of force between charged point-masses, which are even simpler than elasticity.

m. Einstein's geometrodynamics, which looks simple, is interpreted by Sakharov as a correction term in particle physics, which is complicated. Is particle physics, which is complicated, destined some day in its turn to unravel into something simple—something far deeper and far simpler than geometry ("pregeometry"; Chapter 44)?

§17.6. "NO PRIOR GEOMETRY": A FEATURE DISTINGUISHING EINSTEIN'S THEORY FROM OTHER THEORIES OF GRAVITY

Einstein's theory compared with other theories of gravity

Whereas Einstein's theory of gravity is exceedingly compelling, one can readily construct less compelling and less elegant alternative theories. The physics literature is replete with examples [see Ni (1972), and Thorne, Ni, and Will (1971) for reviews]. However, when placed among its competitors, Einstein's theory stands out sharp and clear: it agrees with experiment; most of its competitors do not (Chapters 38–40). It describes gravity entirely in terms of geometry; most of its competitors do not. It is free of any "prior geometry"; most of its competitors are not.

All other theories introduce auxiliary gravitational fields or prior geometry

Set aside, until Chapter 38, the issue of agreement with experiment. Einstein's theory remains unique. Every other theory either introduces auxiliary gravitational fields [e.g., the scalar field of Brans and Dicke (1961)], or involves "prior geometry," or both. Thus, every other theory is more complicated conceptually than Einstein's theory. Every other theory contains elements of complexity for which there is no experimental motivation.

"Prior geometry" defined

Nordstrom's theory as an illustration of prior geometry

The concept of "prior geometry" requires elucidation, not least because the rejection of prior geometry played a key role in the reasoning that originally led Einstein to his geometrodynamic equation $\boldsymbol{G} = 8\pi\boldsymbol{T}$. *By "prior geometry" one means any aspect of the geometry of spacetime that is fixed immutably, i.e., that cannot be changed by changing the distribution of gravitating sources.* Thus, prior geometry is not generated by or affected by matter; it is not dynamic. Example: Nordstrøm (1913) formulated a theory in which the physical metric of spacetime $\boldsymbol{g}$ (the metric that enters into the equivalence principle) is generated by a "background" flat-spacetime metric $\boldsymbol{\eta}$, and by a scalar gravitational field ϕ:

$$\eta^{\alpha\beta}\phi_{,\alpha\beta} = -4\pi\phi\eta^{\alpha\beta}T_{\alpha\beta} \qquad \begin{pmatrix}\text{generation of } \phi \text{ by}\\ \text{stress-energy}\end{pmatrix}, \tag{17.23a}$$

$$g_{\alpha\beta} = \phi^2\eta_{\alpha\beta} \qquad \begin{pmatrix}\text{construction of } \boldsymbol{g}\\ \text{from } \phi \text{ and } \boldsymbol{\eta}\end{pmatrix}. \tag{17.23b}$$

In this theory, the physical metric $\boldsymbol{g}$ (governor of rods and clocks and of test-particle motion) has but one changeable degree of freedom—the freedom in ϕ. The rest of $\boldsymbol{g}$ is fixed by the flat spacetime metric ("prior geometry") $\boldsymbol{\eta}$. One does not remove the prior geometry by rewriting Nordstrøm's equations (17.23) in a form

$$\underbrace{R}_{\begin{bmatrix}\text{curvature scalar}\\ \text{constructed from } \boldsymbol{g}\end{bmatrix}} = 24\,\pi \underbrace{T}_{[g^{\alpha\beta}T_{\alpha\beta}]}, \qquad \underbrace{C^{\alpha\beta}{}_{\mu\nu}}_{\begin{bmatrix}\text{Weyl tensor}\\ \text{constructed from } \boldsymbol{g}\end{bmatrix}} = 0 \tag{17.24}$$

devoid of reference to $\boldsymbol{\eta}$ and ϕ [Einstein and Fokker (1914); exercise 17.8]. Mass can still influence only one degree of freedom in the spacetime geometry. The other degrees of freedom are fixed *à priori*—they are *prior geometry.* And this prior geometry can perfectly well (in principle) be detected by physical experiments that make no reference to any equations (Box 17.3).

Box 17.3 AN EXPERIMENT TO DETECT OR EXCLUDE CERTAIN TYPES OF PRIOR GEOMETRY

(Based on December 1970 discussions between Alfred Schild and Charles W. Misner)

Choose a momentarily static universe populated with a large supply of suitable pulsars. The pulsars should be absolutely regular, periodically emitting characteristic pulses of both gravitational and electromagnetic waves.

Two fleets of spaceships containing receivers are sent out "on station" to collect the experimental data. Admiral Weber's fleet carries gravitational-wave receivers; Admiral Hertz's fleet, electromagnetic receivers. The captain of each spaceship holds himself "on station" by monitoring three suitably chosen pulsars (of identical frequency) and maneuvering so that their pulses always arrive in coincidence. The experimental data he collects consist of the pulses received from all other pulsars, which he is not using for station keeping, each registered as coincident with or interlaced among the reference (stationary) pulses. [For display purposes, the pattern produced by any single pulsar can be converted to acoustic form. The reference pulses can be played acoustically (by the data-processing computer) on one drum at a fixed rate, and the pulses from other pulsars can be played on a second drum. A pattern of rythmic beats will result.]

When the data fleet is checked out and tuned up, each captain reports stationary patterns. Now the experiment begins. One or more massive stars are towed in among the fleet. The fleet reacts to stay on station, and reports changes in the data patterns. The spaceships on the outside edges of the fleet verify that no detectable changes occur at their stations; so the incident radiation from the distant pulsars can be regarded as unaffected by the newly placed stars. Data stations nearer the movable stars report the interesting data.

What are the results?

In a universe governed by the laws of special relativity (spacetime always flat), no patterns change. (Weber's fleet was unable to get checked out in the first place, as no gravitational waves were ever detected from the pulsars). Neither stars, nor anything else, can produce gravitational fields. All aspects of the spacetime geometry are fixed *a priori* (complete prior geometry!). There is no gravity; and no light deflection takes place to make Hertz's captains adjust their positions.

In a universe governed by Nordstrøm's theory of gravity (see text) both fleets get checked out—i.e., both see waves. But neither fleet sees any changes in the rhythmic pattern of beats. The stars being towed about have no influence on either gravitational waves or electromagnetic waves. The prior geometry ($\boldsymbol{\eta}$) present in the theory precludes any light deflection or any gravitational-wave deflection.

In a universe governed by Whitehead's (1922) theory of gravity [see Will (1971b) and references cited therein], radio waves propagate along geodesics of the "physical metric" $\boldsymbol{g}$, and get deflected by the gravitational fields of the stars. But gravitational waves propagate along geodesics of a *flat* background metric $\boldsymbol{\eta}$, and are thus unaffected by the stars. Consequently, Hertz's captains must maneuver to keep on station; and they hear a changing beat pattern between the reference pulsars and the other pulsars. But Weber's fleet remains on station and records no changes in the beat pattern. The prior geometry ($\boldsymbol{\eta}$) shows itself clearly in the experimental result.

In a universe governed by Einstein's theory, both fleets see effects (no sign of prior geometry because Einstein's theory has no prior geometry). Moreover, if the fleets were originally paired, one Weber ship and one Hertz at each station, they remain paired. No differences exist between the propagation of high-frequency light waves and high-frequency gravitational waves. Both propagate along geodesics of $\boldsymbol{g}$.

Mathematics was not sufficiently refined in 1917 to cleave apart the demands for "no prior geometry" and for a "geometric, coordinate-independent formulation of physics." Einstein described both demands by a single phrase, "general covariance." The "no-prior-geometry" demand actually fathered general relativity, but by doing so anonymously, disguised as "general covariance," it also fathered half a century of confusion. [See, e.g., Kretschmann (1917).]

"No prior geometry" as a part of Einstein's principle of "general covariance"

A systematic treatment of the distinction between prior geometry ("absolute objects") and dynamic fields ("dynamic objects") is a notable feature of Anderson's (1967) relativity text.

EXERCISE

Exercise 17.8. EINSTEIN-FOKKER REDUCES TO NORDSTRØM

The vanishing of the Weyl tensor [equation (13.50)] for a spacetime metric $\boldsymbol{g}$ guarantees that the metric is conformally flat—i.e., that there exists a scalar field ϕ such that $\boldsymbol{g} = \phi^2\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a flat-spacetime metric. [See, e.g., Schouten (1954) for proof.] Thus, the Einstein-Fokker equation (17.24), $C^{\alpha\beta}{}_{\mu\nu} = 0$, is equivalent to the Nordstrøm equation (17.23b). With this fact in hand, show that the Einstein-Fokker field equation $R = 24\pi T$ reduces to the Nordstrøm field equation (17.23a).

---

§17.7. A TASTE OF THE HISTORY OF EINSTEIN'S EQUATION

Nothing shows better what an idea is and means today than the battles and changes it has undergone on its way to its present form. A complete history of general relativity would demand a book. Here let a few key quotes from a few of the great papers give a little taste of what a proper history might encompass.

> Einstein (1908): "We . . . will therefore in the following assume the complete physical equivalence of a gravitational field and of a corresponding acceleration of the reference system. . . . the clock at a point P for an observer anywhere in space runs $(1 + \Phi/c^2)$ times faster than the clock at the coordinate origin. . . . it follows that light rays are curved by the gravitational field. . . . an amount of energy E has a mass E/c^2."
>
> Einstein and Grossmann (1913): "The theory described here originates from the conviction that the proportionality between the inertial and the gravitational mass of a body is an exact law of nature that must be expressed as a foundation principle of theoretical physics. . . . An observer enclosed in an elevator has no way to decide whether the elevator is at rest in a static gravitational field or whether the elevator is located in gravitation-free space in an accelerated motion that is maintained by forces acting on the elevator (equivalence hypothesis). . . . In the decay of radium, for example, that decrease [of mass] amounts to 1/10,000 of the total mass. If those changes in inertial mass did not correspond to changes in gravitational mass, then deviations of inertial from gravitational masses would arise that are far larger than the Eötvös experiments allow. It must therefore be considered as very probable that the identity of gravitational and inertial mass is exact.

"The sought for generalization will surely be of the form

$$\Gamma_{\mu\nu} = \kappa T_{\mu\nu},$$

where κ is a constant and $\Gamma_{\mu\nu}$ is a contravariant tensor of the second rank that arises out of the fundamental tensor $g_{\mu\nu}$ through differential operations. . . . it proved impossible to find a differential expression for $\Gamma_{\mu\nu}$ that is a generalization of [Poisson's] $\Delta\phi$, and that is a tensor with respect to arbitrary transformations. . . . It seems most natural to demand that the system of equations should be covariant against arbitrary transformations. That stands in conflict with the result that the equations of the gravitational field do not possess this property."

Einstein and Grossman (1914): "In a 1913 treatment . . . we could not show general covariance for these gravitational equations. [Origin of their difficulty: part of the two-index curvature tensor was put on the left, to constitute the second-order part of the field equation, and part was put on the right with $T_{\mu\nu}$ and was called gravitational stress-energy. It was asked that lefthand and righthand sides transform as tensors, which they cannot do under general coordinate transformations.]

Einstein (1915a): "In recent years I had been trying to found a general theory of relativity on the assumption of the relativity even of nonuniform motions. I believed in fact that I had found the only law of gravitation that corresponds to a reasonably formulated postulate of general relativity, and I sought to establish the necessity of exactly this solution in a paper that appeared last year in these proceedings.

"A renewed analysis showed me that that necessity absolutely was not shown in the approach adopted there; that it nevertheless appeared to be shown rested on an error.

"For these reasons, I lost all confidence in the field equations I had set up, and I sought for an approach that would limit the possibilities in a natural way. In this way I was led back to the demand for the general covariance of the field equations, from which I had departed three years ago, while working with my friend Grossmann, only with a heavy heart. In fact we had already at that time come quite near to the solution of the problem that is given in what follows.

"According to what has been said, it is natural to postulate the field equations of gravitation in the form

$$R_{\mu\nu} = -\kappa T_{\mu\nu},$$

since we already know that these equations are covariant with respect to arbitrary transformations of determinant 1. In fact, these equations satisfy all conditions that we have to impose on them. [Here $R_{\mu\nu}$ is a piece of the Ricci tensor that Einstein regarded as covariant.] . . .

"Equations (22a) give in the first approximation

$$\frac{\partial^2 g^{\alpha\beta}}{\partial x^\alpha \partial x^\beta} = 0.$$

By this [condition] the coordinate system is still not determined, in the sense that for this determination four equations are necessary." (Session of Nov. 4, 1915, published Nov. 11.)

Einstein (1915b): "In a recently published investigation, I have shown how a theory of the gravitational field can be founded on Riemann's covariant theory of many-dimensional manifolds. Here it will now be proved that, by introducing a surely bold additional hypothesis on the structure of matter, a still tighter logical structure of the theory can be achieved. . . . it may very well be possible that in the matter to which the given expression refers, gravitational fields play an essential part. Then $T^{\mu}{}_{\mu}$ can appear to be positive for the entire structure, although in reality only $T^{\mu}{}_{\mu} + t^{\mu}{}_{\mu}$ is positive, and $T^{\mu}{}_{\mu}$ vanishes everywhere. We assume in the following that in fact the condition $T^{\mu}{}_{\mu} = 0$ is fulfilled [quite] generally.

"Whoever does not from the beginning reject the hypothesis that molecular [small-scale] gravitational fields constitute an essential part of matter will see in the following a strong support for this point of view.

"Our hypothesis makes it possible . . . to give the field equations of gravitation in a generally covariant form . . .

$$G_{\mu\nu} = -\kappa T_{\mu\nu}$$

[where $G_{\mu\nu}$ is the Ricci tensor]." (Session of Nov. 11, 1915; published Nov. 18.)

Einstein (1915c): "I have shown that no objection of principle stands in the way of this hypothesis [the field equations], by which space and time are deprived of the last trace of objective reality. In the present work I find an important confirmation of this most radical theory of relativity: it turns out that it explains qualitatively and quantitatively the secular precession of the orbit of Mercury in the direction of the orbital motion, as discovered by Leverrier, which amounts to about 45″ per century, without calling on any special hypothesis whatsoever."

Einstein (1915d; session of Nov. 25, 1915; published Dec. 2): "More recently I have found that one can proceed without hypotheses about the energy tensor of matter when one introduces the energy tensor of matter in a somewhat different way than was done in my two earlier communications. The field equations for the motion of the perihelion of Mercury are undisturbed by this modification. . . .

"Let us put

$$G_{im} = -\kappa\left(T_{im} - \frac{1}{2} g_{im} T\right)$$

[where G_{im} is the Ricci tensor]." . . .

. . . these equations, in contrast to (9), contain no new condition, so that no other assumption has to be made about the energy tensor of matter than obedience to the energy-momentum [conservation] laws.

"With this step, general relativity is finally completed as a logical structure. The postulate of relativity in its most general formulation, which makes the spacetime coordinates into physically meaningless parameters, leads compellingly to a completely determinate theory of gravitation that explains the perihelion motion of Mercury. In contrast, the general-relativity postulate is able to open up to us nothing about the nature of the other processes of nature that special relativity has not already taught. The opinion on this point that I recently expressed in these proceedings was erroneous. Every physical theory compatible with special relativity can be aligned into the system of general relativity by means of the absolute differential calculus, without [general relativity] supplying any criterion for the acceptability of that theory."

Hilbert (1915): "Axiom I [notation changed to conform to usage in this book]. The law of physical events is determined through a world function [Mie's terminology; better known today as "Lagrangian"] L, that contains the following arguments:

$$g_{\mu\nu}, \frac{\partial g_{\mu\nu}}{\partial x^\alpha}, \frac{\partial^2 g_{\mu\nu}}{\partial x^\alpha \, \partial x^\beta},$$

$$A_\sigma, \frac{\partial A_\sigma}{\partial x^\tau},$$

and specifically the variation of the integral

$$\int L(-g)^{1/2}\, d^4x$$

must vanish for [changes in] every one of the 14 potentials $g_{\sigma\nu}$, A_σ. . . .

"Axiom II (axiom of general invariance). The world function L is invariant with respect to arbitrary transformations of the world parameters [coordinates] x^α. . . .

"For the world function L, still further axioms are needed to make its choice unambiguous. If the gravitation equations are to contain only second derivatives of the potentials $g^{\sigma\nu}$, then L must have the form

$$L = R + L_{\text{elec}},$$

where R is the invariant built from the Riemann tensor (curvature of the four-dimensional manifold." (Session of Nov. 20, 1915.)

Einstein (1916c): "Recently H. A. Lorentz and D. Hilbert have succeeded in giving general relativity an especially transparent form in deriving its equations from a single variation principle. This will be done also in the following treatment. There it is my aim to present the basic relations as transparently as possible and in a way as general as general relativity allows."

Einstein (1916b): "From this it follows, first of all, that gravitational fields spread out with the speed of light. . . . [plane] waves transport energy. . . . One thus gets . . . the radiation of the system per unit time. . . .

$$\frac{G}{24\pi} \sum_{\alpha,\beta} \left(\frac{\partial^3 J_{\alpha\beta}}{\partial t^3}\right)^2."$$

Hilbert (1917): "As for the principle of causality, the physical quantities and their time-rates of change may be known at the present time in any given coordinate system; a prediction will then have a physical meaning only when it is invariant with respect to all those transformations for which exactly those coordinates used for the present time remain unchanged. I declare that predictions of this kind for the future are all uniquely determined; that is, that the causality principle holds in this formulation:

"From the knowledge of the 14 physical potentials $g_{\mu\nu}$, A_σ, in the present, all predictions about the same quantities in the future follow necessarily and uniquely insofar as they have physical meaning."

CHAPTER 18

WEAK GRAVITATIONAL FIELDS

The way that can be walked on is not the perfect way.
The word that can be said is not the perfect word.

LAO-TZU (~3rd century B.C.)

§18.1. THE LINEARIZED THEORY OF GRAVITY

"Linearized theory of gravity":

(1) as weak-field limit of general relativity

Because of the geometric language and abbreviations used in writing them, Einstein's field equations, $G_{\mu\nu} = 8\pi T_{\mu\nu}$, hardly seem to be differential equations at all, much less ones with many familiar properties. The best way to see that they are is to apply them to weak-field situations

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1, \tag{18.1}$$

e.g., to the solar system, where $|h_{\mu\nu}| \sim |\Phi| \lesssim M_\odot/R_\odot \sim 10^{-6}$; or to a weak gravitational wave propagating through interstellar space.

(2) as standard "field-theory" description of gravity in "flat spacetime"

In a weak-field situation, one can expand the field equations in powers of $h_{\mu\nu}$, using a coordinate frame where (18.1) holds; and without much loss of accuracy, one can keep only linear terms. The resulting formalism is often called *"the linearized theory of gravity,"* because it is an important theory in its own right. In fact, it is precisely this "linearized theory" that one obtains when one asks for the classical field corresponding to quantum-mechanical particles of (1) zero rest mass and (2) spin two in (3) flat spacetime [see Fierz and Pauli (1939)]. Track-2 readers have already explored linearized theory somewhat in §7.1, exercise 7.3, and Box 7.1. There it went under the alternative name, "tensor-field theory of gravity in flat spacetime."

(3) as a foundation for "deriving" general relativity

Just as one can "descend" from general relativity to linearized theory by linearizing about flat spacetime (see below), so one can "bootstrap" one's way back up from linearized theory to general relativity by imposing consistency between the linearized field equations and the equations of motion. or, equivalently, by asking about: (1) the stress-energy carried by the linearized gravitational field $h_{\mu\nu}$; (2) the influence of this stress-energy acting as a source for corrections $h^{(1)}{}_{\mu\nu}$ to the field; (3) the stress-energy carried by the corrections $h^{(1)}{}_{\mu\nu}$; (4) the influence of this stress-energy acting as a source for corrections $h^{(2)}{}_{\mu\nu}$ to the corrections $h^{(1)}{}_{\mu\nu}$; (5) the stress-energy carried by the corrections to the corrections; and so on. This alternative way to derive general relativity has been developed and explored by Gupta (1954, 1957, 1962), Kraichnan (1955), Thirring (1961), Feynman (1963a), Weinberg (1965), and Deser (1970). But because the outlook is far from geometric (see Box 18.1), the details of the derivation are not presented here. (But see part 5 of Box 17.2.)

Details of linearized theory:

Here attention focuses on deriving linearized theory from general relativity. Adopt the form (18.1) for the metric components. The resulting connection coefficients [equations (8.24b)], when linearized in the metric perturbation $h_{\mu\nu}$, read

(1) connection coefficients

$$\begin{aligned}\Gamma^{\mu}{}_{\alpha\beta} &= \frac{1}{2}\eta^{\mu\nu}(h_{\alpha\nu,\beta} + h_{\beta\nu,\alpha} - h_{\alpha\beta,\nu}) \\ &\equiv \frac{1}{2}(h_{\alpha}{}^{\mu}{}_{,\beta} + h_{\beta}{}^{\mu}{}_{,\alpha} - h_{\alpha\beta}{}^{,\mu}).\end{aligned} \tag{18.2}$$

The second line here introduces the convention, used routinely whenever one expands in powers of $h_{\mu\nu}$, that indices of $h_{\mu\nu}$ are raised and lowered using $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$, *not* $g^{\mu\nu}$ and $g_{\mu\nu}$. A similar linearization of the Ricci tensor [equation (8.47)] yields

$$\begin{aligned}R_{\mu\nu} &= \Gamma^{\alpha}{}_{\mu\nu,\alpha} - \Gamma^{\alpha}{}_{\mu\alpha,\nu} \\ &= \frac{1}{2}(h_{\mu}{}^{\alpha}{}_{,\nu\alpha} + h_{\nu}{}^{\alpha}{}_{,\mu\alpha} - h_{\mu\nu,\alpha}{}^{\alpha} - h_{,\mu\nu}),\end{aligned} \tag{18.3}$$

where

$$h \equiv h^{\alpha}{}_{\alpha} = \eta^{\alpha\beta}h_{\alpha\beta}. \tag{18.4}$$

After a further contraction to form $R \equiv g^{\mu\nu}R_{\mu\nu} \approx \eta^{\mu\nu}R_{\mu\nu}$, one finds that the Einstein equations, $2G_{\mu\nu} = 16\pi T_{\mu\nu}$, read

$$\begin{aligned}h_{\mu\alpha,\nu}{}^{\alpha} + h_{\nu\alpha,\mu}{}^{\alpha} - h_{\mu\nu,\alpha}{}^{\alpha} - h_{,\mu\nu} \\ - \eta_{\mu\nu}(h_{\alpha\beta}{}^{,\alpha\beta} - h_{,\beta}{}^{\beta}) = 16\pi T_{\mu\nu}.\end{aligned} \tag{18.5}$$

The number of terms has increased in passing from $R_{\mu\nu}$ (18.3) to $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ (18.5), but this annoyance can be counteracted by defining

(2) "gravitational potentials" $\bar{h}_{\mu\nu}$

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h \tag{18.6}$$

and using a bar to imply a corresponding operation on any other symmetric tensor.

Box 18.1 DERIVATIONS OF GENERAL RELATIVITY FROM GEOMETRIC VIEWPOINT AND FROM SPIN-TWO VIEWPOINT, COMPARED AND CONTRASTED

	Einstein derivation	*Spin-2 derivation*
Nature of primordial spacetime geometry?	Not primordial; geometry is a dynamic participant in physics	"God-given" flat Lorentz spacetime manifold
Topology (multiple connectedness) of spacetime?	Laws of physics are local; they do not specify the topology	Simply connected Euclidean topology
Vision of physics?	Dynamic geometry is the "master field" of physics	This field, that field, and the other field all execute their dynamics in a flat-spacetime manifold
Starting points for this derivation of general relativity	1. Equivalence principle (world lines of photons and test particles are geodesics of the spacetime geometry) 2. That tensorial conserved quantity which is derived from the curvature (Cartan's moment of rotation) is to be identified with the tensor of stress-momentum-energy (see Chapter 15).	1. Begin with field of spin two and zero rest mass in flat spacetime. 2. Stress-energy tensor built from this field serves as a source for this field.
Resulting equations	Einstein's field equations	Einstein's field equations
Resulting assessment of the spacetime geometry from which derivation started	Fundamental dynamic participant in physics	None. Resulting theory eradicates original flat geometry from all equations, showing it to be unobservable
View about the greatest single crisis of physics to emerge from these equations: complete gravitational collapse	Central to understanding the nature of matter and the universe	Unimportant or at most peripheral

Thus $G_{\mu\nu} = \bar{R}_{\mu\nu}$ to first order in the $h_{\mu\nu}$, and $\bar{\bar{h}}_{\mu\nu} = h_{\mu\nu}$; i.e., $h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h}$. With this notation *the linearized field equations become*

$$-\bar{h}_{\mu\nu,\alpha}{}^{\alpha} - \eta_{\mu\nu}\bar{h}_{\alpha\beta,}{}^{\alpha\beta} + \bar{h}_{\mu\alpha,}{}^{\alpha}{}_{\nu} + \bar{h}_{\nu\alpha,}{}^{\alpha}{}_{\mu} = 16\pi T_{\mu\nu}. \tag{18.7}$$

(3) linearized field equations

The first term in these linearized equations is the usual flat-space d'Alembertian, and the other terms serve merely to keep the equations "gauge-invariant" (see Box

18.2). In Box 18.2 it is shown that, without loss of generality, one can impose the "gauge conditions"

(4) gauge conditions

$$\bar{h}^{\mu\alpha}{}_{,\alpha} = 0. \tag{18.8a}$$

These gauge conditions are the tensor analog of the Lorentz gauge $A^{\alpha}{}_{,\alpha} = 0$ of electromagnetic theory. The field equations (18.7) then become

(5) field equations and metric in Lorentz gauge

$$-\bar{h}_{\mu\nu,\alpha}{}^{\alpha} = 16\pi T_{\mu\nu}. \tag{18.8b}$$

The gauge conditions (18.8a), the field equations (18.8b), and the definition of the metric

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} = \eta_{\mu\nu} + \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h} \tag{18.8c}$$

are the fundamental equations of the linearized theory of gravity written in Lorentz gauge.

EXERCISES

Exercise 18.1. GAUGE INVARIANCE OF THE RIEMANN CURVATURE

Show that in linearized theory the components of the Riemann tensor are

$$R_{\alpha\mu\beta\nu} = \frac{1}{2}(h_{\alpha\nu,\mu\beta} + h_{\mu\beta,\nu\alpha} - h_{\mu\nu,\alpha\beta} - h_{\alpha\beta,\mu\nu}). \tag{18.9}$$

Then show that these components are left unchanged by a gauge transformation of the form discussed in Box 18.2 [equation (4b)]. Since the Einstein tensor is a contraction of the Riemann tensor, this shows that it is also gauge-invariant.

Exercise 18.2. JUSTIFICATION OF LORENTZ GAUGE

Let a particular solution to the field equations (18.7) of linearized theory be given, in an arbitrary gauge. Show that there necessarily exist four generating functions $\xi_\mu(t, x^j)$ whose gauge transformation [Box 18.2, eq. (4b)] makes

$$\bar{h}^{\text{new}\,\mu\alpha}{}_{,\alpha} = 0 \qquad \text{(Lorentz gauge)}.$$

Also show that a subsequent gauge transformation leaves this Lorentz gauge condition unaffected if and only if its generating functions satisfy the sourceless wave equation

$$\xi^{\alpha,\beta}{}_{\beta} = 0.$$

Exercise 18.3. EXTERNAL FIELD OF A STATIC, SPHERICAL BODY

Consider the external gravitational field of a static spherical body, as described in the body's (nearly) Lorentz frame—i.e., in a nearly rectangular coordinate system $|h_{\mu\nu}| \ll 1$, in which the body is located at $x = y = z = 0$ for all t. By fiat, adopt Lorentz gauge.

(a) Show that the field equations (18.8b) and gauge conditions (18.8a) imply

$$\bar{h}_{00} = 4M/(x^2 + y^2 + z^2)^{1/2}, \qquad \bar{h}_{0j} = \bar{h}_{jk} = 0,$$

$$h_{00} = h_{xx} = h_{yy} = h_{zz} = 2M/(x^2 + y^2 + z^2)^{1/2}, \qquad h_{\alpha\beta} = 0 \text{ if } \alpha \neq \beta,$$

where M is a constant (the mass of the body; see §19.3).

Box 18.2 GAUGE TRANSFORMATIONS AND COORDINATE TRANSFORMATIONS IN LINEARIZED THEORY

A. The Basic Equations of Linearized Theory, written in any coordinate system that is nearly globally Lorentz, are (18.1) and (18.7):

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1; \tag{1}$$

$$-\bar{h}_{\mu\nu,\alpha}{}^{\alpha} - \eta_{\mu\nu}\bar{h}_{\alpha\beta}{}^{,\alpha\beta} + \bar{h}_{\mu\alpha}{}^{,\alpha}{}_{\nu} + \bar{h}_{\nu\alpha}{}^{,\alpha}{}_{\mu} = 16\pi T_{\mu\nu}. \tag{2}$$

Two different types of coordinate transformations connect nearly globally Lorentz systems to each other: global Lorentz transformations, and infinitesimal coordinate transformations.

1. *Global Lorentz Transformations:*

$$x^{\mu} = \Lambda^{\mu}{}_{\alpha'}x^{\alpha'}, \qquad \Lambda^{\mu}{}_{\alpha'}\Lambda^{\nu}{}_{\beta'}\eta_{\mu\nu} = \eta_{\alpha'\beta'}. \tag{3a}$$

These transform the metric coefficients via

$$\eta_{\alpha'\beta'} + h_{\alpha'\beta'} = g_{\alpha'\beta'} = \frac{\partial x^{\mu}}{\partial x^{\alpha'}}\frac{\partial x^{\nu}}{\partial x^{\beta'}}g_{\mu\nu} = \Lambda^{\mu}{}_{\alpha'}\Lambda^{\nu}{}_{\beta'}(\eta_{\mu\nu} + h_{\mu\nu})$$
$$= \eta_{\alpha'\beta'} + \Lambda^{\mu}{}_{\alpha'}\,\Lambda^{\nu}{}_{\beta'}h_{\mu\nu}.$$

Thus, $h_{\mu\nu}$—and likewise $\bar{h}_{\mu\nu}$—transform like components of a tensor in flat spacetime

$$h_{\alpha'\beta'} = \Lambda^{\mu}{}_{\alpha'}\Lambda^{\nu}{}_{\beta'}h_{\mu\nu}. \tag{3b}$$

2. *Infinitesimal Coordinate Transformations* (creation of "ripples" in the coordinate system):

$$x^{\mu'}(\mathcal{P}) = x^{\mu}(\mathcal{P}) + \xi^{\mu}(\mathcal{P}), \tag{4a}$$

where $\xi^{\mu}(\mathcal{P})$ are four arbitrary functions small enough to leave $|h_{\mu'\nu'}| \ll 1$. Infinitesimal transformations of this sort make tiny changes in the functional forms of all scalar, vector, and tensor fields. *Example:* the temperature T is a unique function of position, $T(\mathcal{P})$; so when written as a function of coordinates it changes

$$T(x^{\mu'} = a^{\mu}) = T(x^{\mu} + \xi^{\mu} = a^{\mu}) = T(x^{\mu} = a^{\mu} - \xi^{\mu})$$
$$= T(x^{\mu} = a^{\mu}) - T_{,\mu}\xi^{\mu};$$

i.e., if $\xi^0 = 0.001 \sin(x^1)$, and if $T = \cos^2(x^0)$, then

$$T = \cos^2(x^{0'}) + 0.002 \sin(x^{1'}) \cos(x^{0'}) \sin(x^{0'}).$$

Box 18.2 (*continued*)

These tiny changes can be ignored in all quantities except the metric, where tiny deviations from $\eta_{\mu\nu}$ contain all the information about gravity. The usual tensor transformation law for the metric

$$g_{\rho'\sigma'}[x^{\alpha'}(\mathscr{P})] = g_{\mu\nu}[x^{\alpha}(\mathscr{P})] \frac{\partial x^{\mu}}{\partial x^{\rho'}} \frac{\partial x^{\nu}}{\partial x^{\rho'}},$$

when combined with the transformation law (4a) and with

$$g_{\mu\nu}[x^{\alpha}(\mathscr{P})] = \eta_{\mu\nu} + h_{\mu\nu}[x^{\alpha}(\mathscr{P})],$$

reveals that

$$\begin{aligned} g_{\rho'\sigma'}(x^{\alpha'} = a^{\alpha}) = \eta_{\rho\sigma} + h_{\rho\sigma}(x^{\alpha} = a^{\alpha}) - \xi_{\rho,\sigma} - \xi_{\sigma,\rho} \\ + \text{negligible corrections} \sim h_{\rho\sigma,\alpha}\xi^{\alpha} \text{ and } \sim h_{\rho\alpha}\xi^{\alpha}{}_{,\sigma}. \end{aligned}$$

Hence, *the metric perturbation functions in the new ($x^{\mu'}$) and old (x^{μ}) coordinate systems are related by*

$$h_{\mu\nu}^{\text{new}} = h_{\mu\nu}^{\text{old}} - \xi_{\mu,\nu} - \xi_{\nu,\mu}, \tag{4b}$$

whereas the functional forms of all other scalars, vectors, and tensors are unaltered, to within the precision of linearized theory.

B. Gauge Transformations and Gauge Invariance. In linearized theory one usually regards equation (4b) as gauge transformations, analogous to those

$$A_{\mu}^{\text{new}} = A_{\mu}^{\text{old}} + \Psi_{,\mu} \tag{5a}$$

of electromagnetic theory. The fact that gravitational gauge transformations do not affect the functional forms of scalars, vectors, or tensors (i.e., observables) is called "gauge invariance." Just as a straightforward calculation reveals the gauge invariance of the electromagnetic field,

$$F_{\mu\nu}^{\text{new}} = A_{\nu,\mu}^{\text{new}} - A_{\mu,\nu}^{\text{new}} = A_{\nu,\mu}^{\text{old}} + \Psi_{,\nu\mu} - A_{\mu,\nu}^{\text{old}} - \Psi_{,\mu\nu} = F_{\mu\nu}^{\text{old}}, \tag{5b}$$

so a straightforward calculation (exercise 18.1) reveals the gauge invariance of the Riemann tensor ·

$$R_{\mu\nu\alpha\beta}^{\text{new}} = R_{\mu\nu\alpha\beta}^{\text{old}}. \tag{6}$$

Such gauge invariance was already guaranteed by the fact that $R_{\mu\nu\alpha\beta}$ are the components of a tensor, and are thus essentially the same whether calculated in an orthonormal frame $g_{\hat{\mu}\hat{\nu}} = \eta_{\mu\nu}$, in the old coordinates where $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}^{\text{old}}$, or in the new coordinates where $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}^{\text{new}}$.

Like the Riemann tensor, the Einstein tensor and the stress-energy tensor are unaffected by gauge transformations. Hence, *if one knows a specific solution* $\bar{h}_{\mu\nu}$ *to the linearized field equations* (2) *for a given* $T^{\mu\nu}$, *one can obtain another solution that describes precisely the same physical situation* (*all observables unchanged*) *by the change of gauge* (4), *in which* ξ_μ *are four arbitrary but small functions.*

C. Lorentz Gauge. One can show (exercise 18.2) that for any physical situation, one can specialize the gauge (i.e., the coordinates) so that $\bar{h}^{\mu\alpha}{}_{,\alpha} = 0$. This is the Lorentz gauge introduced in §18.1. The Lorentz gauge is not fixed uniquely. The gauge condition $\bar{h}^{\mu\alpha}{}_{,\alpha} = 0$ is left unaffected by any gauge transformation for which

$$\xi^{\alpha,\beta}{}_{\beta} = 0.$$

(See exercise 18.2.)

D. Curvilinear Coordinate Systems. Once the gauge has been fixed by fiat for a given system (e.g., the solar system), one can regard $h_{\mu\nu}$ and $\bar{h}_{\mu\nu}$ as components of tensors in flat spacetime; and one can regard the field equations (2) and the chosen gauge conditions as geometric, coordinate-independent equations in flat spacetime. This viewpoint allows one to use curvilinear coordinates (e.g., spherical coordinates centered on the sun), if one wishes. But in doing so, one must everywhere replace the Lorentz components of the metric, $\eta_{\mu\nu}$, by the metric's components $g_{\mu\nu\,\mathrm{flat}}$ in the flat-spacetime curvilinear coordinate system; and one must replace all ordinary derivatives ("commas") in the field equations and gauge conditions by covariant derivatives whose connection coefficients come from $g_{\mu\nu\,\mathrm{flat}}$. See exercise 18.3 for an example.

(b) Adopt spherical polar coordinates,

$$x = r \sin\theta\cos\phi, \qquad y = r\sin\theta\sin\phi, \qquad z = r\cos\theta.$$

By regarding $h_{\mu\nu}$ and $\bar{h}_{\mu\nu}$ as components of tensors in flat spacetime (see end of Box 18.2), and by using the usual tensor transformation laws, put the solution found in (a) into the form

$$\bar{h}_{00} = 4M/r, \qquad \bar{h}_{0j} = \bar{h}_{jk} = 0,$$

$$h_{00} = \frac{2M}{r}, \qquad h_{0j} = 0, \qquad h_{jk} = \frac{2M}{r} g_{jk\,\mathrm{flat}}$$

where $g_{\alpha\beta\,\mathrm{flat}}$ are the components of the flat-spacetime metric in the spherical coordinate system

$$g_{00\,\mathrm{flat}} = -1, \qquad g_{rr\,\mathrm{flat}} = 1, \qquad g_{\theta\theta\,\mathrm{flat}} = r^2,$$
$$g_{\phi\phi\,\mathrm{flat}} = r^2\sin^2\theta, \qquad g_{\alpha\beta\,\mathrm{flat}} = 0 \text{ when } \alpha \neq \beta.$$

Thereby conclude that the general relativistic line element, accurate to linearized order, is

$$ds^2 = -(1 - 2M/r)\, dt^2 + (1 + 2M/r)(dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2).$$

(c) Derive this general, static, spherically symmetric, Lorentz-gauge, vacuum solution to the linearized field equations from scratch, working entirely in spherical coordinates. [*Hint:* As discussed at the end of Box 18.2, $\eta_{\mu\nu}$ in equation (18.8c) must be replaced by $g_{\mu\nu\,\mathrm{flat}}$; and in the field equations and gauge conditions (18.8a, b), all commas (partial derivatives) must be replaced by covariant derivatives, whose connection coefficients come from $g_{\mu\nu\,\mathrm{flat}}$.]
(d) Calculate the Riemann curvature tensor for this gravitational field. The answer should agree with equation (1.14).

---

§18.2. GRAVITATIONAL WAVES

Linearized theory and electromagnetic theory compared

The gauge conditions and field equations (18.8a, b) of linearized theory bear a close resemblance to the equations of electromagnetic theory in Lorentz gauge and flat spacetime,

$$A^{\alpha}{}_{,\alpha} = 0, \tag{18.10a}$$

$$-A^{\mu}{}_{,\alpha}{}^{\alpha} = 4\pi J^{\mu}. \tag{18.10b}$$

They differ only in the added index ($h^{\mu\nu}$ versus A^{μ}, $T^{\mu\nu}$ versus J^{μ}). Consequently, from past experience with electromagnetic theory, one can infer much about linearized gravitation theory.

For example, the field equations (18.8b) must have gravitational-wave solutions. The analog of the electromagnetic plane wave

$$A^x = A^x(t - z), \qquad A^y = A^y(t - z), \qquad A^z = 0, \qquad A^0 = 0,$$

Plane gravitational waves

will be the gravitational plane wave

$$\begin{gathered} \bar{h}^{xx} = \bar{h}^{xx}(t - z), \qquad \bar{h}^{xy} = \bar{h}^{xy}(t - z), \qquad \bar{h}^{yy} = \bar{h}^{yy}(t - z), \\ \bar{h}^{\mu 0} = \bar{h}^{\mu z} = 0 \text{ for all } \mu. \end{gathered} \tag{18.11}$$

Although a detailed study of such waves will be delayed until Chapters 35–37, some properties of these waves are explored in the exercises at the end of the next section.

§18.3. EFFECT OF GRAVITY ON MATTER

How to analyze effects of weak gravity on matter

The effects of weak gravitational fields on matter can be computed by using the linearized metric (18.1) and Christoffel symbols (18.2) in the appropriate equations of motion—i.e., in the geodesic equation (for the motion of particles or light rays), in the hydrodynamic equations (for fluid matter), in Maxwell's equations (for electromagnetic waves), or in the equation $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ for the total stress-energy tensor of whatever fields and matter may be present. Exercises 18.5, 18.6 and 18.7 provide examples, as do the Newtonian-limit calculations in exercises 16.1 and 16.4, and in §17.4. If, however, the lowest-order (linearized) gravitational "forces" (Christoffel-symbol terms) have a significant influence on the motion of the sources of the gravitational field, one finds that the linearized field equation (18.7) is inadequate, and better approximations to Einstein's equations must be considered. [Thus emission of gravitational waves by a mechanically or electrically driven oscillator falls within the scope of linearized theory, but emission by a double-star system, or by stellar oscillations that gravitational forces maintain, will require discussion of nonlinear terms (gravitational "stress-energy") in the Einstein equations; see §§36.9 to 36.11.]

The above conclusions follow from a consideration of conservation laws associated with the linearized field equation. Just as the electromagnetic equations (18.10a, b) guarantee charge conservation

$$J^{\mu}{}_{,\mu} = 0, \qquad \int_{\text{all space}} J^0(t, \boldsymbol{x})\, dx\, dy\, dz \equiv Q = \text{const},$$

Conservation of 4-momentum and angular momentum in linearized theory

so the gravitational equations (18.8a, b) guarantee conservation of the total 4-momentum and angular momentum of any body bounded by vacuum:

$$T^{\mu\nu}{}_{,\nu} = 0, \tag{18.12a}$$

$$\int_{\text{body}} T^{\mu 0}(t, \boldsymbol{x})\, dx\, dy\, dz \equiv P^{\mu} = \text{const}; \tag{18.12b}$$

$$(x^{\alpha} T^{\beta\mu} - x^{\beta} T^{\alpha\mu})_{,\mu} = 0, \tag{18.13a}$$

$$\int_{\text{body}} (x^{\alpha} T^{\beta 0} - x^{\beta} T^{\alpha 0})\, dx\, dy\, dz \equiv J^{\alpha\beta} = \text{const}. \tag{18.13b}$$

(See §5.11 for the basic properties of angular momentum in special relativity. The angular momentum here is calculated relative to the origin of the coordinate system.) Now it is important that the stress-energy components $T^{\mu\nu}$, which appear in the linearized field equations (18.7) and in these conservation laws, are precisely the components one would calculate using special relativity (with $g_{\mu\nu} = \eta_{\mu\nu}$). As a result, the energy-momentum conservation formulated here contains no contributions or effects of gravity! From this one sees that linearized theory assumes that gravitational forces do no significant work. For example, energy losses due to gravitational radiation-damping forces are neglected by linearized theory. Similarly, conservation of 4-momentum P^{μ} for each of the bodies acting as sources of $h_{\mu\nu}$ means that each body moves along a geodesic of $\eta_{\mu\nu}$ (straight lines in the nearly Lorentz coordinate system) rather than along a geodesic of $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. Thus, linearized theory can be used to calculate the motion of test particles and fields, using $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$; but to include gravitational corrections to the motion of the sources themselves—to allow them to satisfy $T^{\mu\nu}{}_{;\nu} = 0$ rather than $T^{\mu\nu}{}_{,\nu} = 0$—one must reinsert into the field equations the nonlinear terms that linearized theory discards. (See, e.g., Chapter 20 on conservation laws; §§36.9–36.11 on the generation of gravitational waves and radiation reaction; and Chapter 39 on the post-Newtonian approximation.)

Limit on validity of linearized theory: gravity must not affect motions of sources significantly

The energy, momentum, and angular momentum radiated by gravitational waves in linearized theory can be calculated by special-relativistic methods analogous to those used in electromagnetic theory for electromagnetic waves [Fiertz and Pauli (1939)], but it will be more informative and powerful to use a fully gravitational approach (Chapters 35 and 36).

EXERCISES

Exercise 18.4. SPACETIME CURVATURE FOR A PLANE GRAVITATIONAL WAVE

Calculate the components of the Riemann curvature tensor [equations (18.9)] for the gravitational plane wave (18.11). [*Answer:*

$$R_{x0x0} = -R_{y0y0} = -R_{x0xz} = +R_{y0yz} = +R_{xzxz} = -R_{yzyz} = -\frac{1}{4}(\bar{h}_{xx} - \bar{h}_{yy})_{,tt};$$

$$R_{x0y0} = -R_{x0yz} = +R_{xzyz} = -R_{xzy0} = -\frac{1}{2}\bar{h}_{xy,tt};$$

all other components vanish except those obtainable from the above by the symmetries $R_{\alpha\beta\gamma\delta} = R_{[\alpha\beta][\gamma\delta]} = R_{\gamma\delta\alpha\beta}$.

Exercise 18.5. A PRIMITIVE GRAVITATIONAL-WAVE DETECTOR (see Figure 18.1)

Two beads slide almost freely on a smooth stick; only slight friction impedes their sliding. The stick falls freely through spacetime, with its center moving along a geodesic and its ends attached to gyroscopes, so they do not rotate. The beads are positioned equidistant (distance $\frac{1}{2}\ell$) from the stick's center. Plane gravitational waves [equation (18.11) and exercise 18.4], impinging on the stick, push the beads back and forth ("geodesic deviation"; "tidal gravitational forces"). The resultant friction of beads on stick heats the stick; and the passage of the waves is detected by measuring the rise in stick temperature.* (Of course, this is not the best of all conceivable designs!) Neglecting the effect of friction on the beads' motion, calculate the proper distance separating them as a function of time. [*Hints:* Let $\boldsymbol{\xi}$ be the separation between the beads; and let $\boldsymbol{n} = \boldsymbol{\xi}/|\boldsymbol{\xi}|$ be a unit vector that points along the stick in the stick's own rest frame. Then their separation has magnitude $\ell = \boldsymbol{\xi}\cdot\boldsymbol{n}$. The fact that the stick is nonrotating is embodied in a parallel-transport law for $\boldsymbol{n}$, $\nabla_{\boldsymbol{u}}\boldsymbol{n} = 0$. ("Fermi-Walker transport" of §§6.5, 6.6, and 13.6 reduces to parallel transport, because the stick moves along a geodesic with $\boldsymbol{a} = \nabla_{\boldsymbol{u}}\boldsymbol{u} = 0$.) Thus,

$$d\ell/d\tau = \nabla_{\boldsymbol{u}}(\boldsymbol{\xi}\cdot\boldsymbol{n}) = (\nabla_{\boldsymbol{u}}\boldsymbol{\xi})\cdot\boldsymbol{n},$$
$$d^2\ell/d\tau^2 = \nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}(\boldsymbol{\xi}\cdot\boldsymbol{n}) = (\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{\xi})\cdot\boldsymbol{n},$$

where τ is the stick's proper time. But $\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{\xi}$ is produced by the Riemann curvature of the wave (geodesic deviation):

$$\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{\xi} = \text{projection along } \boldsymbol{n} \text{ of } [-\boldsymbol{Riemann}\,(\ldots, \boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{u})].$$

(The geodesic-deviation forces perpendicular to the stick, i.e., perpendicular to $\boldsymbol{n}$, are coun-

---

*This thought experiment was devised by Bondi [1957, 1965; Bondi and McCrea (1960)] as a means for convincing skeptics of the reality of gravitational waves.

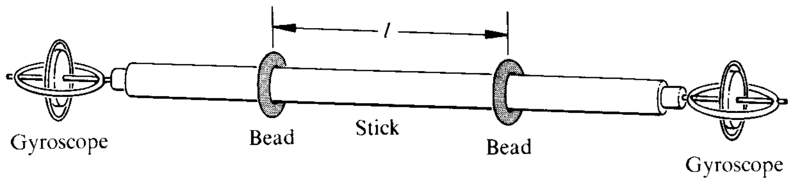


Figure 18.1.
A primitive detector for gravitational waves, consisting of a beaded stick with gyroscopes on its ends [Bondi (1957)]. See exercise 18.5 for discussion.

terbalanced by the stick's pushing back on the beads to stop them from passing through it—no penetration of matter by matter!) Thus,

$$d^2\ell/d\tau^2 = -\textbf{Riemann}\,(\ldots, \boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{u})\cdot \boldsymbol{n} = -\textbf{Riemann}\,(\boldsymbol{n}, \boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{u}).$$

Evaluate this acceleration in the stick's local Lorentz frame. Orient the coordinates so the waves propagate in the z-direction and the stick's direction has components $n^z = \cos\theta$, $n^x = \sin\theta\cos\phi$, $n^y = \sin\theta\sin\phi$. Solve the resulting differential equation for $\ell(\tau)$.] [*Answer:*

$$\ell = \ell_0\left[1 + \frac{1}{4}(\bar{h}_{xx} - \bar{h}_{yy})\sin^2\theta\cos 2\phi + \frac{1}{2}\bar{h}_{xy}\sin^2\theta\sin 2\phi\right],$$

where $\bar{h}_{jk}$ are evaluated on the stick's world line ($x = y = z = 0$). Notice that, if the stick is oriented along the direction of wave propagation (if $\theta = 0$), the beads do not move. In this sense, the effect of the waves (geodesic deviation) is purely transverse. For further discussion, see §§35.4 to 35.6.]

---

§18.4. NEARLY NEWTONIAN GRAVITATIONAL FIELDS

The general solution to the linearized field equations in Lorentz gauge [equations (18.8a, b)] lends itself to expression as a retarded integral of the form familiar from electromagnetic theory:

Retarded-integral solution of linearized field equation

$$\bar{h}_{\mu\nu}(t, \boldsymbol{x}) = \int \frac{4T_{\mu\nu}(t - |\boldsymbol{x} - \boldsymbol{x}'|, \boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|}\, d^3x'. \tag{18.14}$$

The gravitational-wave aspects of this solution will be studied in Chapter 36. Here focus attention on a nearly Newtonian source: $T_{00} \gg |T_{0j}|$, $T_{00} \gg |T_{jk}|$, and velocities slow enough that retardation is negligible. In this case, (18.14) reduces to

Newtonian gravity as a limit of linearized theory

$$\bar{h}_{00} = -4\Phi, \qquad \bar{h}_{0j} = \bar{h}_{jk} = 0, \tag{18.15a}$$

$$\Phi(t, \boldsymbol{x}) = -\int \frac{T_{00}(t, \boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|}\, d^3x' = \text{Newtonian potential}. \tag{18.15b}$$

The corresponding metric (18.8c) is

$$\begin{aligned} ds^2 &= -(1 + 2\Phi)\,dt^2 + (1 - 2\Phi)(dx^2 + dy^2 + dz^2) \qquad (18.15c)\\ &\approx -(1 - 2M/r)\,dt^2 + (1 + 2M/r)(dx^2 + dy^2 + dz^2) \text{ far from source.} \end{aligned}$$

The errors in this metric are: (1) missing corrections of order Φ^2 due to nonlinearities of which linearized theory is oblivious; (2) missing corrections due to setting $\bar{h}_{0j} = 0$ (these are of order $\bar{h}_{0j} \sim \Phi v$, where $v \sim |T_{0j}|/T_{00}$ is a typical velocity in the source); (3) missing corrections due to setting $\bar{h}_{jk} = 0$ [these are of order $\bar{h}_{jk} \sim \Phi(|T_{jk}|/T_{00})$]. In the solar system all these errors are $\sim 10^{-12}$, whereas $\Phi \sim 10^{-6}$.

Bending of light and gravitational redshift predicted by linearized theory

Passive correspondence with Newtonian theory demanded only that $g_{00} = -(1 + 2\Phi)$; see equation (17.19). However, linearized theory determines all the metric coefficients, up to errors of $\sim\Phi v$, $\sim\Phi^2$, and $\sim\Phi(|T_{jk}|/T_{00})$. This is sufficient accuracy to predict correctly (fractional errors $\sim 10^{-6}$) the bending of light and the gravitational redshift in the solar system, but not perihelion shifts.

EXERCISES

Exercise 18.6. BENDING OF LIGHT BY THE SUN

To high precision, the sun is static and spherical, so its external line element is (18.15c) with $\Phi = -M/r$; i.e.,

$$ds^2 = -(1 - 2M/r)\,dt^2 + (1 + 2M/r)(dx^2 + dy^2 + dz^2) \text{ everywhere outside sun.} \tag{18.16}$$

A photon moving in the equatorial plane ($z = 0$) of this curved spacetime gets deflected very slightly from the world line

$$x = t, \qquad y = b \equiv \text{“impact parameter,”} \qquad z = 0. \tag{18.17}$$

Calculate the amount of deflection as follows.

(a) Write down the geodesic equation (16.4a) for the photon's world line,

$$\frac{dp^\alpha}{d\lambda^*} + \Gamma^\alpha{}_{\beta\gamma}p^\beta p^\gamma = 0. \tag{18.18}$$

[Here $\boldsymbol{p} = d/d\lambda^* =$ (4-momentum of photon) $=$ (tangent vector to photon's null geodesic).]

(b) By evaluating the connection coefficients in the equatorial plane, and by using the approximate values, $|p^y| \ll p^0 \approx p^x$, of the 4-momentum components corresponding to the approximate world line (18.17), show that

$$\frac{dp^y}{d\lambda^*} = \frac{-2Mb}{(x^2 + b^2)^{3/2}}p^x\frac{dx}{d\lambda^*}, \qquad p^x = p^0\left[1 + O\left(\frac{M}{b}\right)\right] = \text{const}\left[1 + O\left(\frac{M}{b}\right)\right].$$

(c) Integrate this equation for p^y, assuming $p^y = 0$ at $x = -\infty$ (photon moving precisely in x-direction initially); thereby obtain

$$p^y(x = +\infty) = -\frac{4M}{b}p^x.$$

(d) Show that this corresponds to deflection of light through the angle

$$\Delta\phi = 4M/b = 1''.75\,(R_\odot/b), \tag{18.19}$$

where $R_\odot$ is the radius of the sun. For a comparison of this prediction with experiment, see Box 40.1.

Exercise 18.7. GRAVITATIONAL REDSHIFT

(a) Use the geodesic equation for a photon, written in the form

$$dp_\mu/d\lambda^* - \Gamma^\alpha{}_{\mu\beta}p_\alpha p^\beta = 0,$$

to prove that any photon moving freely in the sun's gravitational field [line element (18.16)] has $dp_0/d\lambda^* = 0$; i.e.,

$$p_0 = \text{constant along photon's world line.} \tag{18.20}$$

(b) An atom at rest on the sun's surface emits a photon of wavelength λ_e, as seen in its orthonormal frame. [Note:

$$h\nu_e = h/\lambda_e = (\text{energy atom measures}) = -\boldsymbol{p}\cdot\boldsymbol{u}_e, \tag{18.21}$$

where $\boldsymbol{p}$ is the photon's 4-momentum and $\boldsymbol{u}_e$ is the emitter's 4-velocity.] An atom at rest far from the sun receives the photon, and measures its wavelength to be λ_r. [Note: $h/\lambda_r = -\boldsymbol{p}\cdot\boldsymbol{u}_r$.] Show that the photon is redshifted by the amount

$$z \equiv \frac{\lambda_r - \lambda_e}{\lambda_e} = \frac{M_\odot}{R_\odot} = 2\times 10^{-6}. \tag{18.22}$$

[*Hint:* $\boldsymbol{u}_r = \partial/\partial t$; $\boldsymbol{u}_e = (1-2M/r)^{-1/2}\,\partial/\partial t$. Why?] For further discussion of the gravitational redshift and experimental results, see §§7.4 and 38.5; also Figures 38.1 and 38.2.

CHAPTER 19

MASS AND ANGULAR MOMENTUM OF A GRAVITATING SYSTEM

§19.1. EXTERNAL FIELD OF A WEAKLY GRAVITATING SOURCE

Metric far from a weakly gravitating system, as a power series in $1/r$:

Consider an isolated system with gravity so weak that in calculating its structure and motion one can completely ignore self-gravitational effects. (This is true of an asteroid, and of a nebula with high-energy electrons and protons spiraling in a magnetic field; it is not true of the Earth or the sun.) Assume nothing else about the system—for example, by contrast with Newtonian theory, allow velocities to be arbitrarily close to the speed of light, and allow stresses T^{jk} and momentum densities T^{0j} to be comparable to the mass-energy density T^{00}.

(1) derivation

Calculate the weak gravitational field,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{19.1}$$

$$\bar{\bar{h}}_{\mu\nu} \equiv h_{\mu\nu} = \int \frac{4\bar{T}_{\mu\nu}(t - |\boldsymbol{x} - \boldsymbol{x}'|, \boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|} d^3x', \tag{19.2}$$

produced by such a system [see "barred" version of equation (18.14)]. Restrict attention to the spacetime region far outside the system, and expand $h_{\mu\nu}$ in powers of $\boldsymbol{x}'/r \equiv \boldsymbol{x}'/|\boldsymbol{x}|$, using the relations

$$\bar{T}_{\mu\nu}(t - |\boldsymbol{x} - \boldsymbol{x}'|, \boldsymbol{x}') = \sum_{n=0}^{\infty} \frac{1}{n!}\left[\frac{\partial^n}{\partial t^n}\bar{T}_{\mu\nu}(t - r, \boldsymbol{x}')\right](r - |\boldsymbol{x} - \boldsymbol{x}'|)^n, \tag{19.3a}$$

$$r - |\boldsymbol{x} - \boldsymbol{x}'| = x^j\left(\frac{x^{j'}}{r}\right) + \frac{1}{2}\frac{x^j x^k}{r}\left(\frac{x^{j'}x^{k'} - r'^2\,\delta_{jk}}{r^2}\right) + \cdots, \tag{19.3b}$$

$$\frac{1}{|\boldsymbol{x} - \boldsymbol{x}'|} = \frac{1}{r} + \frac{x^j}{r^2}\frac{x^{j'}}{r} + \frac{1}{2}\frac{x^j x^k}{r^3}\frac{(3x^{j'}x^{k'} - r'^2\,\delta_{jk})}{r^2} + \cdots. \tag{19.3c}$$

Perform the calculation in the system's rest frame, where

$$P^j \equiv \int T^{0j}\, d^3x = 0, \tag{19.4a}$$

with origin of coordinates at the system's center of mass

$$\int x^j T^{00}\, d^3x = 0. \tag{19.4b}$$

The result, after a change of gauge to simplify h_{00} and h_{0j}, is

(2) result

$$\begin{aligned} ds^2 = &-\left[1 - \frac{2M}{r} + 0\left(\frac{1}{r^3}\right)\right] dt^2 - \left[4\epsilon_{jk\ell} S^k \frac{x^\ell}{r^3} + 0\left(\frac{1}{r^3}\right)\right] dt\, dx^j \\ &+ \left[\left(1 + \frac{2M}{r}\right)\delta_{jk} + \begin{pmatrix}\text{gravitational radiation terms} \\ \text{that die out as } 0(1/r)\end{pmatrix}\right] dx^j\, dx^k. \end{aligned} \tag{19.5}$$

(see exercise 19.1 for derivation.) Here M and S^k are the body's mass and intrinsic angular momentum.

$$M = \int T^{00}\, d^3x, \tag{19.6a}$$

$$S_k = \int \epsilon_{k\ell m} x^\ell T^{m0}\, d^3x. \tag{19.6b}$$

The corresponding Newtonian potential is

$$\Phi = -\frac{1}{2}(g_{00} - \eta_{00}) = -\frac{M}{r} + 0\left(\frac{1}{r^3}\right). \tag{19.6c}$$

How metric depends on system's mass M and angular momentum S

Conclusion: With an appropriate choice of gauge, Φ and g_{00} far from any weak source are time-independent and are determined uniquely by the source's mass M; g_{0j} is time-independent and is fixed by the source's intrinsic angular momentum S^j; but g_{jk} has time-dependent terms (gravitational waves!) of $0(1/r)$.

The rest of this chapter focuses on the "imprints" of the mass and angular momentum in the gravitational field; the gravitational waves will be ignored, or almost so, until Chapter 35.

EXERCISE

Exercise 19.1. DERIVATION OF METRIC FAR OUTSIDE A WEAKLY GRAVITATING BODY

(a) Derive equation (19.5). [*Hints:* (1) Follow the procedure outlined in the text. (2) When calculating h_{00}, write out explicitly the $n = 0$ and $n = 1$ terms of (19.2), to precision $0(1/r^2)$, and simplify the $n = 0$ term using the identities

$$T^{jk} = \frac{1}{2}(T^{00}x^j x^k)_{,00} + (T^{\ell j}x^k + T^{\ell k}x^j)_{,\ell} - \frac{1}{2}(T^{\ell m}x^j x^k)_{,\ell m}, \tag{19.7a}$$

$$T^{\ell\ell}x^m = \left(T^{0\ell}x^\ell x^m - \frac{1}{2}T^{0m}r^2\right)_{,0} + \left(T^{\ell k}x^k x^m - \frac{1}{2}T^{\ell m}r^2\right)_{,\ell}. \tag{19.7b}$$

(Verify that these identities follow from $T^{\alpha\beta}{}_{,\beta} = 0$.) (3) When calculating h_{0m}, write out explicitly the $n = 0$ term of (19.2), to precision $0(1/r^2)$, and simplify it using the identity

$$T^{0k}x^j + T^{0j}x^k = (T^{00}x^jx^k)_{,0} + (T^{0\ell}x^jx^k)_{,\ell}. \tag{19.7c}$$

(Verify that this follows from $T^{\alpha\beta}{}_{,\beta} = 0$.) (4) Simplify h_{00} and h_{0m} by the gauge transformation generated by

$$\begin{aligned}\xi_0 = {} & \frac{1}{2r}\frac{\partial}{\partial t}\int T^{00\prime}r'^2\,d^3x' + \frac{x^j}{r^3}\int\left(T^{0k\prime}x^{k\prime}x^{j\prime} - \frac{1}{2}T^{0j\prime}r'^2\right)d^3x' \\ & + \int(T_{00}' + T_{\ell\ell}')\left[\frac{x^jx^{j\prime}}{r^2} + \frac{(3x^{j\prime}x^{k\prime} - r'^2\delta_{jk})x^jx^k}{2r^4}\right]d^3x' \\ & + \sum_{n=2}^{\infty}\frac{1}{n!}\frac{\partial^{n-1}}{\partial t^{n-1}}\int(T_{00}' + T_{kk}')\frac{(r - |\boldsymbol{x} - \boldsymbol{x}'|)^n}{|\boldsymbol{x} - \boldsymbol{x}'|}\,d^3x', \\ \xi_m = {} & -\frac{2x^j}{r^3}\int T_{00}'x^{j\prime}x^{m\prime}\,d^3x' + 4\sum_{n=1}^{\infty}\frac{1}{n!}\frac{\partial^{n-1}}{\partial t^{n-1}}\int T_{0m}'\frac{(r - |\boldsymbol{x} - \boldsymbol{x}'|)^n}{|\boldsymbol{x} - \boldsymbol{x}'|}\,d^3x' \\ & + \frac{x^m}{r}\xi_0 - \frac{1}{2}\left(\frac{1}{r}\right)_{,m}\int T_{00}'r'^2\,d^3x' - \left(\frac{x^k}{r^2}\right)_{,m}\int\left(T^{0j\prime}x^{j\prime}x^{k\prime} - \frac{1}{2}T^{0k\prime}r'^2\right)d^3x' \\ & - \sum_{n=2}^{\infty}\frac{1}{n!}\frac{\partial^{n-2}}{\partial t^{n-2}}\int(T_{00}' + T_{kk}')\left[\frac{(r - |\boldsymbol{x} - \boldsymbol{x}'|)^n}{|\boldsymbol{x} - \boldsymbol{x}'|}\right]_{,m}d^3x'\end{aligned}$$

Here $T_{\mu\nu}'$ denotes $T_{\mu\nu}(t - r, \boldsymbol{x}')$.]

(b) Prove that the system's mass and angular momentum are conserved. [*Note:* Because $T^{\alpha\beta}{}_{,\beta} = 0$ (self-gravity has negligible influence), the proof is no different here than in flat spacetime (Chapter 5).]

---

§19.2. MEASUREMENT OF THE MASS AND ANGULAR MOMENTUM

For a weakly gravitating system:

The values of a system's mass and angular momentum can be measured by probing the imprint they leave in its external gravitational field. Of all tools one might use to probe, the simplest is a test particle in a gravitationally bound orbit. If the particle is sufficiently far from the source, its motion is affected hardly at all by the source's angular momentum or by the gravitational waves; only the spherical, Newtonian part of the gravitational field has a significant influence. Hence, the particle moves in an elliptical Keplerian orbit. To determine the source's mass M, one need only apply Kepler's third law (perhaps better called "Kepler's 1-2-3 law"):

(1) total mass M can be measured by applying Kepler's "1-2-3" law to orbiting particles

$$M = \left(\frac{2\pi}{\text{orbital period}}\right)^2\left(\begin{matrix}\text{Semi-major axis}\\ \text{of ellipse}\end{matrix}\right)^3; \qquad \text{i.e., } M^1 = \omega^2a^3. \tag{19.8}$$

The source's angular momentum is not measured quite so easily. One must use a probe that is insensitive to Newtonian gravitational effects, but "feels" the off-diagonal term,

$$g_{0j} = -2\epsilon^{jk\ell}S^kx^\ell/r^3, \tag{19.9}$$

in the metric (19.5). One such probe is the precession of the perihelion of a corevolving satellite, relative to the precession for a counterrevolving satellite. A gyroscope is another such probe. Place a gyroscope at rest in the source's gravitational field. By a force applied to its center of mass, prevent it from falling. As time passes, the g_{0j} term in the metric will force the gyroscope to precess relative to the basis vectors $\partial/\partial x^j$; and since these basis vectors are "tied" to the coordinate system, which in turn is tied to the Lorentz frames at infinity, which in turn are tied to the "fixed stars" (cf. §39.12), the precession is relative to the "fixed stars." The angular velocity of precession, as derived in exercise 19.2, is

(2) total angular momentum $\boldsymbol{S}$ can be measured by examining the precession of gyroscopes

$$\boldsymbol{\Omega} = \frac{1}{r^3}\left[-\boldsymbol{S} + \frac{3(\boldsymbol{S}\cdot\boldsymbol{x})\boldsymbol{x}}{r^2}\right]. \tag{19.10}$$

One sometimes says that the source's rotation "drags the inertial frames near the source," thereby forcing the gyroscope to precess. For further discussion, see §§21.12, 40.7, and 33.4.

EXERCISE

Exercise 19.2. GYROSCOPE PRECESSION

Derive equation (19.10) for the angular velocity of gyroscope precession. [*Hints:* Place an orthonormal tetrad at the gyroscope's center of mass. Tie the tetrad rigidly to the coordinate system, and hence to the "fixed stars"; more particularly, choose the tetrad to be that basis $\{e_{\hat\alpha}\}$ which is dual to the following 1-form basis:

$$\boldsymbol{\omega}^{\hat t} = [1-(2M/r)]^{1/2}\,\boldsymbol{d}t + 2\epsilon_{jkl}S^k(x^l/r^3)\,\boldsymbol{d}x^j, \qquad \boldsymbol{\omega}^{\hat j} = [1+(2M/r)]^{1/2}\,\boldsymbol{d}x^j. \tag{19.11}$$

The spatial legs of the tetrad, $\boldsymbol{e}_{\hat j}$, rotate relative to the gyroscope with an angular velocity $\boldsymbol{\omega}$ given by [see equation (13.69)]

$$-\epsilon_{\hat i\hat j\hat k}\,\omega^{\hat k} = \Gamma_{\hat i\hat j\hat 0}.$$

Consequently, the gyroscope's angular momentum vector $\boldsymbol{L}$ precesses relative to the tetrad with angular velocity $\boldsymbol{\Omega} = -\boldsymbol{\omega}$:

$$\frac{dL^{\hat j}}{d\hat t} = \epsilon_{\hat j\hat k\hat l}\,\Omega^{\hat k}L^{\hat l}, \qquad \epsilon_{\hat i\hat j\hat k}\,\Omega^{\hat k} = \Gamma_{\hat i\hat j\hat 0}. \tag{19.12}$$

Calculate $\Gamma_{\hat i\hat j\hat 0}$ for the given orthonormal frame, and thereby obtain equation (19.10) for $\boldsymbol{\Omega}$.]

---

§19.3. MASS AND ANGULAR MOMENTUM OF FULLY RELATIVISTIC SOURCES

Abandon, now, the restriction to weakly gravitating sources. Consider an isolated, gravitating system inside which spacetime may or may not be highly curved—a black hole, a neutron star, the Sun, . . . But refuse, for now, to analyze the system's interior or the "strong-field region" near the system. Instead, restrict attention to the weak gravitational field far from the source, and analyze it using linearized theory in vacuum. Expand $h_{\mu\nu}$ in multipole moments and powers of $1/r$; and adjust the gauge, the Lorentz frame, and the origin of coordinates to simplify the resulting metric. The outcome of such a calculation is a gravitational field identical to that for a weak source [equation (19.5)]! (Details of the calculation are not spelled out here because of their length; but see exercise 19.3.)

But before accepting this as the distant field of an arbitrary source, one should examine the nonlinear effects in the vacuum field equations. Two types of nonlinearities turn out to be important far from the source: (1) nonlinearities in the static, Newtonian part of the metric, which generate metric corrections

$$\delta g_{00} = -2M^2/r^2, \qquad \delta g_{jk} = \frac{3}{2}(M^2/r^2)\,\delta_{jk},$$

(see exercise 19.3 and §39.8), thereby putting the metric into the form

Metric far from *any* gravitating system, as a power series in $1/r$

$$\begin{aligned} ds^2 = &-\left[1 - \frac{2M}{r} + \frac{2M^2}{r^2} + 0\left(\frac{1}{r^3}\right)\right]dt^2 - \left[4\epsilon_{jk\ell}S^k\frac{x^\ell}{r^3} + 0\left(\frac{1}{r^3}\right)\right]dt\,dx^j \\ &+\left[\left(1 + \frac{2M}{r} + \frac{3M^2}{2r^2}\right)\delta_{jk} + \begin{pmatrix}\text{gravitational radiation terms}\\ \text{that die out as } 0(1/r)\end{pmatrix}\right]dx^j\,dx^k; \end{aligned} \tag{19.13}$$

(2) a gradual decrease in the source's mass, gradual changes in its angular momentum, and gradual changes in its "rest frame" to compensate for the mass, angular momentum, and linear momentum carried off by gravitational waves (see Box 19.1, which is best read only after finishing this section).

By measuring the distant spacetime geometry (19.13) of a given source, one cannot discover whether that source has strong internal gravity, or weak. But when one expresses the constants M and S_j, which determine g_{00} and g_{0j}, as integrals over the interior of the source, one discovers a crucial difference: if the internal gravity is weak, then linearized theory is valid throughout the source, and

Failure of volume integrals for M and S when source has strong internal gravity

$$M = \int T_{00}\,d^3x, \qquad S_j = \int \epsilon_{jk\ell}x^k T^{\ell 0}\,d^3x; \tag{19.14}$$

but if the gravity is strong, these formulas fail. Does this failure prevent one, for strong gravity, from identifying the constants M and S_j of the metric (19.13) as the source's mass and angular momentum? Not at all, according to the following argument.

Consider, first, the mass of the sun. For the sun one expects Newtonian theory to be highly accurate (fractional errors $\sim M_\odot/R_\odot \sim 10^{-6}$); so one can assert that the constant M appearing in the line element (19.13) is, indeed

$$M = \int \rho\,d^3x = \int T_{00}\,d^3x = \text{total mass}.$$

But might this assertion be wrong? To gain greater confidence and insight, adopt the viewpoint of "*controlled ignorance*"; i.e., do not pretend to know more than what is needed. (This style of physical argument goes back to Newton's famous "Hypotheses non fingo," i.e. "I do not feign hypotheses.") In evaluating the volume integral of T_{00} (usual Newtonian definition of M), one needs a theory of the internal structure of the sun. For example, one must know that the visible surface layers of the sun do not hide a massive central core, so dense and large that relativistic gravitational fields $|\Phi| \sim 1$ exist there. If one makes use in the analysis of a fluid-type stress-energy tensor $T^{\mu\nu}$, one needs to know equations of state, opacities, and theories of energy generation and transport. One needs to justify the fluid description as an adequate approximation to the atomic constitution of matter. One needs to assume that an ultimate theory of matter explaining the rest masses of protons and electrons will not assign an important fraction of this mass to strong (nonlinear) gravitational fields on a submicroscopic scale. It is plausible that one could do all this, but it is also obvious that this is not the way the mass of the sun is, in fact, determined by astronomers! Theories of stellar structure are adjusted to give the observed mass; they are not constructed to let one deduce the mass from nongravitational observations. The mass of the sun is measured in practice by studying the orbits of planets in its external gravitational field, a procedure equivalent to reading the mass M off the line element (19.13), rather than evaluating the volume integral $\int T^{00}\, d^3x$.

Definition of "total mass-energy" M and "angular momentum" S in terms of external gravitational field

To avoid all the above uncertainties, and to make theory correspond as closely as possible to experiment, *one defines the "total mass-energy" M of the sun or any other body to be the constant that appears in the line element (19.13) for its distant external spacetime geometry. Similarly, one defines the body's intrinsic angular momentum as the constant 3-vector $\boldsymbol{S}$ appearing in its line element (19.13).* Operationally, the total mass-energy M is measured via Kepler's third law; the angular momentum $\boldsymbol{S}$ is measured via its influence on the precession of a gyroscope or a planetary orbit. This is as true when the body is a black hole or a neutron star as when it is the sun.

$\boldsymbol{S}$ as a geometric object in an asymptotically flat region far outside source

What kind of a geometric object is the intrinsic angular momentum $\boldsymbol{S}$? It is defined by measurements made far from the source, where, with receding distance, spacetime is becoming flatter and flatter (asymptotically flat). Thus, it can be regarded as a 3-vector in the *"asymptotically flat spacetime"* that surrounds the source. But in what Lorentz frame is $\boldsymbol{S}$ a 3-vector? Clearly, in the asymptotic Lorentz frame where the line element (19.13) is valid; i.e., in the asymptotic Lorentz frame where the source's distant "coulomb" ("M/r") field is static; i.e., in the *"asymptotic rest frame"* of the source. Alternatively, one can regard $\boldsymbol{S}$ as a 4-vector, $\boldsymbol{S}$, which is purely spatial ($S^0 = 0$) in the asymptotic rest frame. If one denotes the 4-velocity of the asymptotic rest frame by $\boldsymbol{U}$, then the fact that $\boldsymbol{S}$ is purely spatial can be restated geometrically as $\boldsymbol{S} \cdot \boldsymbol{U} = 0$, or

"Asymptotic rest frame" and "total 4-momentum"

$$\boldsymbol{S} \cdot \boldsymbol{P} = 0, \tag{19.15}$$

where

$$\boldsymbol{P} \equiv M\boldsymbol{U} \equiv \text{"total 4-momentum of source"} \tag{19.16}$$

is still another vector residing in the asymptotically flat region of spacetime.

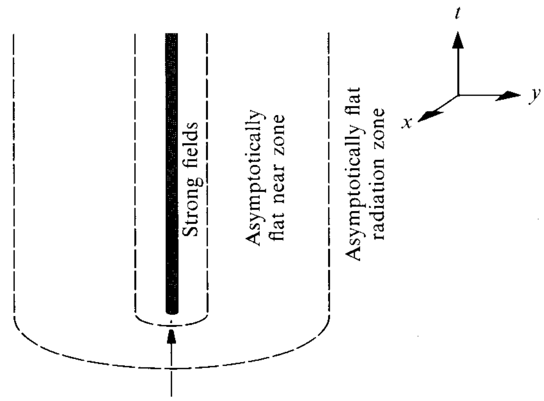
Conservation laws for total 4-momentum and angular momentum

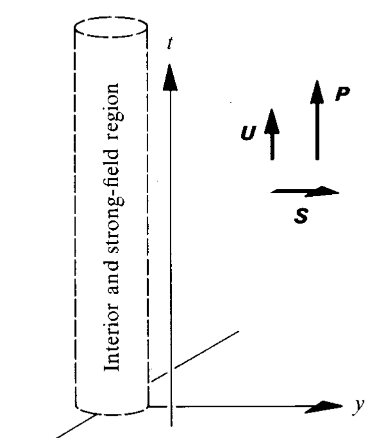
The total 4-momentum $\boldsymbol{P}$ and intrinsic angular momentum $\boldsymbol{S}$ satisfy conservation laws that are summarized in Box 19.1. These conservation laws are valuable tools in gravitation theory and relativistic astrophysics, but the derivation of these laws (Chapter 20) does not compare in priority to topics such as neutron stars and basic cosmology; so most readers will wish to skip it on a first reading of this book.

(*continued on page 456*)

Box 19.1 TOTAL MASS-ENERGY, 4-MOMENTUM, AND ANGULAR MOMENTUM OF AN ISOLATED SYSTEM

A. Spacetime is divided into (1) the source's interior; which is surrounded by (2) a strong-field vacuum region; which in turn is surrounded by (3) a weak-field, asymptotically flat, near-zone region; which in turn is surrounded by (4) a weak-field, asymptotically flat, radiation-zone region. This box and this chapter treat only the asymptotically flat regions. The interior and strong-field regions are treated in the next chapter.





Asymptotic rest frame

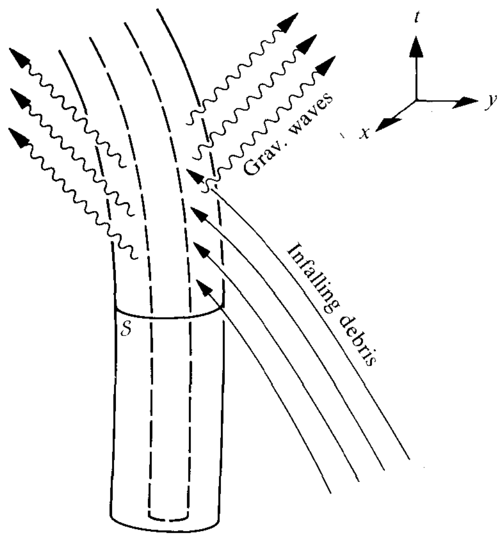
B. The asymptotic rest frame of the source is that global, asymptotically Lorentz frame (coordinates t, x, y, z) in which the distant, "coulomb" part of the source's field is at rest (see diagram). The asymptotic rest frame does not extend into the strong-field region; any such extension of it would necessarily be forced by the curvature into a highly non-Lorentz, curvilinear form. The spatial origin of the asymptotic rest frame is so adjusted that the source is centered on it—i.e., that the distant Newtonian potential is $\Phi = -M/(x^2 + y^2 + z^2)^{1/2} + 0(1/r^3)$; i.e., that Φ has no dipole term, $\boldsymbol{D}\cdot\boldsymbol{x}/r^3$, such as would originate from an offset of the coordinates.

C. To the source one can attribute a total mass-energy M, a 4-velocity $\boldsymbol{U}$, a total 4-momentum $\boldsymbol{P}$, and an intrinsic angular momentum vector, $\boldsymbol{S}$. The 4-vectors $\boldsymbol{U}$, $\boldsymbol{P}$, and $\boldsymbol{S}$ reside in the asymptotically flat region of spacetime and can be moved about freely there (negligible curvature $\Longrightarrow$ parallel transport around closed curves does not change $\boldsymbol{U}$, $\boldsymbol{P}$, or $\boldsymbol{S}$). The source's 4-velocity $\boldsymbol{U}$ is defined to equal the 4-velocity of the asymptotic rest frame ($U^0 = 1$, $\boldsymbol{U} = 0$ in rest frame). The total mass-energy M is measured via Kepler's third ("1-2-3") law [equation (19.8)]. The total 4-momentum is defined by $\boldsymbol{P} \equiv M\boldsymbol{U}$. The intrinsic angular momentum $\boldsymbol{S}$ is orthogonal to the 4-velocity $\boldsymbol{U}$, $\boldsymbol{S}\cdot\boldsymbol{U} = 0$ (so $S^0 = 0$; $\boldsymbol{S} \neq 0$ in general in asymptotic rest frame); $\boldsymbol{S}$ is measured via gyroscope precession or differential perihelion precession (§19.2).

In the asymptotic rest frame, with an appropriate choice of gauge (i.e., of ripples in the coordinates), *the slight deviations from flat-spacetime geometry are described by the line element*

$$
\begin{aligned}
ds^2 = &-\left[1 - \frac{2M}{r} + \frac{2M^2}{r^2} + 0\left(\frac{1}{r^3}\right)\right] dt^2 - \left[4\epsilon_{jkl} S^k \frac{x^l}{r^3} + 0\left(\frac{1}{r^3}\right)\right] dt\, dx^j \\
&+ \left[\left(1 + \frac{2M}{r} + \frac{3M^2}{2r^2}\right)\delta_{jk} + (\text{gravitational radiation terms})\right] dx^j\, dx^k.
\end{aligned} \tag{1}
$$

D. Conservation of 4-momentum and angular momentum: Suppose that particles fall into a source or are ejected from it; suppose that electromagnetic waves flow in and out; suppose the source emits gravitational waves. All such processes break the source's isolation and can change its total 4-momentum $\boldsymbol{P}$, its intrinsic angular momentum $\boldsymbol{S}$, and its asymptotic rest frame. Surround the source with a spherical shell $\mathcal{S}$, which is far enough out to be in the asymptotically flat region. Keep this shell always at rest in the source's momentary asymptotic rest frame. By probing the source's gravitational field near $\mathcal{S}$, measure its 4-momentum $\boldsymbol{P}$ and intrinsic angular momentum $\boldsymbol{S}$ as functions of the shell's proper time τ. An analysis given in the next chapter reveals that the 4-momentum is conserved, in the sense that



Interstellar debris falls into a black hole, and gravitational waves emerge.

$$
\frac{dP^\alpha}{d\tau} = -\int_{\mathcal{S}} T^{\alpha j} n_j\, d(\text{area}) = \begin{pmatrix}\text{rate at which 4-momentum} \\ \text{flows inward through shell}\end{pmatrix}, \tag{2}
$$

where $\boldsymbol{n}$ is the unit outward normal to $\mathcal{S}$ and the integral is performed in the shell's momentary rest frame. In words: *the rate at which 4-momentum flows through the shell, as measured in the standard special relativistic manner, equals the rate of change of the source's gravitationally measured 4-momentum. Similarly, the angular momentum is conserved in the sense that*

$$
\frac{dS_i}{d\tau} = -\int_{\mathcal{S}} (\epsilon_{ijk} x^j T^{kl}) n_l\, d(\text{area}) = \begin{pmatrix}\text{rate at which angular} \\ \text{momentum flows inward} \\ \text{through the shell}\end{pmatrix}, \tag{3a}
$$

$$
\frac{dS_0}{d\tau} = -\frac{dU^\alpha}{d\tau} S_\alpha = \begin{pmatrix}\text{change required to keep } \boldsymbol{S} \text{ orthogonal to } \boldsymbol{U}; \\ \text{“Fermi-Walker-transport law”; cf. §§6.5, 13.6}\end{pmatrix}. \tag{3b}
$$

In these conservation laws $T^{\alpha\beta}$ is the total stress-energy tensor at the shell, including contributions from matter, electromagnetic fields, and gravitational waves. The gravitational-wave contribution, called $T^{(GW)\alpha\beta}$, is treated in Chapter 35.

Note: The conservation laws in the form stated above contain fractional errors of order M/r (contributions from "gravitational potential energy" of infalling material), but such errors go to zero in the limit of a very large shell ($r \longrightarrow \infty$).

Note: The formulation of these conservation laws given in the next chapter is more precise and more rigorous, but less physically enlightening than the one here.

EXERCISE

Exercise 19.3. GRAVITATIONAL FIELD FAR FROM A STATIONARY, FULLY RELATIVISTIC SOURCE

Derive the line element (19.13) for the special case of a source that is time-independent ($g_{\mu\nu,t} = 0$). This can be a difficult problem, if one does not proceed carefully along the following outlined route. (1) Initially ignore all nonlinearities in the Einstein field equations. The field is weak far from the source. These nonlinearities will be absent from the dominant terms. (2) Calculate the dominant terms using linearized theory in the Lorentz gauge [equations (18.8)]. (3) In particular, write the general solution to the vacuum, time-independent wave equation (18.8b) in the following form involving $n^j \equiv x^j/r \equiv$ (unit vector in radial direction):

$$\begin{aligned}
\bar{h}_{00} &= \frac{A^0}{r} + \frac{B^j n^j}{r^2} + 0\left(\frac{1}{r^3}\right),\\
\bar{h}_{0j} &= \frac{A^j}{r} + \frac{B^{jk} n^k}{r^2} + 0\left(\frac{1}{r^3}\right),\\
\bar{h}_{jk} &= \frac{A^{jk}}{r} + \frac{B^{jk\ell} n^\ell}{r^2} + 0\left(\frac{1}{r^3}\right),\\
A^{jk} &= A^{(jk)}, \qquad B^{jk\ell} = B^{(jk)\ell}.
\end{aligned} \tag{19.17}$$

(Round brackets denote symmetrization.) (4) Then impose the Lorentz gauge conditions $\bar{h}_\alpha{}^\beta{}_{,\beta} = 0$ on this general solution, thereby learning

$$\begin{aligned}
A^j = 0, &\qquad A^{jk} = 0,\\
B^{jk}(\delta^{jk} &- 3n^j n^k) = 0,\\
B^{jk\ell}(\delta^{k\ell} &- 3n^k n^\ell) = 0.
\end{aligned} \tag{19.18}$$

(5) Write B^{jk} as the sum of its trace $3B$, its traceless symmetric part S^{jk}, and its traceless antisymmetric part (these are its "irreducible parts"):

$$B^{jk} = B\delta^{jk} + S^{jk} + \epsilon^{jk\ell}F^\ell, \qquad S^{jj} = 0. \tag{19.19}$$

Show that any tensor B^{jk} can be put into such a form. Then show that the gauge conditions (19.18) imply $S^{jk} = 0$. (6) Similarly show that any tensor $B^{jk\ell}$ that is symmetric on its first two indices can be put into the form

$$\begin{aligned}
&B^{jk\ell} = \delta^{jk}A^\ell + C^{(j}\delta^{k)\ell} + \epsilon^{m\ell(j}E^{k)m} + S^{jk\ell},\\
&E^{km} \text{ symmetric and traceless, i.e., } \quad E^{km} = E^{(km)}, \qquad E^{kk} = 0,\\
&S^{jk\ell} \text{ symmetric and traceless, i.e., } \qquad S^{jk\ell} = S^{(jk\ell)},\\
&\qquad\qquad S^{jj\ell} = S^{jkk} = S^{jkj} = 0.
\end{aligned} \tag{19.20}$$

Then show that the gauge conditions (19.18) imply $C^j = -2A^j$ and $E^{km} = S^{jk\ell} = 0$. (7) Combining all these results, conclude that

$$\begin{aligned}
\bar{h}_{00} &= \frac{A^0}{r} + \frac{B^j n^j}{r^2} + 0\left(\frac{1}{r^3}\right),\\
\bar{h}_{0j} &= \frac{\epsilon^{jk\ell} n^k F^\ell}{r^2} + \frac{B n^j}{r^2} + 0\left(\frac{1}{r^3}\right),\\
\bar{h}_{jk} &= \frac{\delta^{jk}A^\ell n^\ell - A^j n^k - A^k n^j}{r^2} + 0\left(\frac{1}{r^3}\right).
\end{aligned} \tag{19.21}$$

Then use gauge transformations, which stay within Lorentz gauge, to eliminate B and A^j from $\bar{h}_{0j}$ and $\bar{h}_{jk}$; so

$$\begin{aligned}\bar{h}_{00} &= \frac{A^0}{r} + \frac{(B^j + A^j)n^j}{r^2} + 0\left(\frac{1}{r^3}\right),\\ \bar{h}_{0j} &= \frac{\epsilon^{jk\ell} n^k F^\ell}{r^2} + 0\left(\frac{1}{r^3}\right),\\ \bar{h}_{jk} &= 0\left(\frac{1}{r^3}\right).\end{aligned} \tag{19.22}$$

(8) Translate the origin of coordinates so $x^j{}_{\text{new}} = x^j{}_{\text{old}} - (B^j + A^j)/A^0$; in the new coordinate system $\bar{h}_{\alpha\beta}$ has the same form as (19.22), but with $B^j + A^j$ removed. From the resultant $\bar{h}_{\alpha\beta}$, construct the metric and redefine the constants A^0 and F^ℓ to agree with equation (19.13). (9) All linear terms in the metric are now accounted for. The dominant nonlinear terms must be proportional to the square, $(M/r)^2$, of the dominant linear term. The easiest way to get the proportionality constant is to take the Schwarzschild geometry for a fully relativistic, static, spherical source [equation (31.1)], by a change of coordinates put it in the form

$$ds^2 = -\left(\frac{1 - M/2r}{1 + M/2r}\right)^2 dt^2 + \left(1 + \frac{M}{2r}\right)^4 (dx^2 + dy^2 + dz^2) \tag{19.23}$$

(exercise 25.8), and expand it in powers of M/r.

§19.4. MASS AND ANGULAR MOMENTUM OF A CLOSED UNIVERSE

"There are no snakes in Ireland."

Statement of St. Patrick after driving the snakes out of Ireland (legend*)

There is no such thing as "the energy (or angular momentum, or charge) of a closed universe," according to general relativity, and this for a simple reason. To weigh something one needs a platform on which to stand to do the weighing.

For a closed universe the total mass-energy M and angular momentum $\boldsymbol{S}$ are undefined and undefinable

To weigh the sun, one measures the periods and semimajor axes of planetary orbits, and applies Kepler's "1-2-3" law, $M = \omega^2 a^3$. To measure the angular momentum, $\boldsymbol{S}$, of the sun (a task for space technology in the 1970's or 1980's!), one measures the precession of a gyroscope in a near orbit about the sun, or one examines some other aspect of the "dragging of inertial frames." To determine the electric charge

*Stokes (1887) and other standard references deny this legend. In part I of Stokes the basic manuscript references are listed, including especially codex manuscript Rawlinson B.512 in 154 folios, in double columns, written by various hands in the fourteenth and fifteenth centuries (*cf. Catalogi codicum manuscriptorum Bibliothecae Bodleianae Partis Quintae Fasciculus Primus*, Oxford, 1862, col. 728–732). In this manuscript, folio 97b.1, line 14, reads in the translation of Stokes, Part I, p. xxx: "as Paradise is without beasts, without a snake, without a lion, without a dragon, without a scorpion, without a mouse, without a frog, so is Ireland in the same manner without any harmful animal, save only the wolf. . ."

of a body, one surrounds it by a large sphere, evaluates the electric field normal to the surface at each point on this sphere, integrates over the sphere, and applies the theorem of Gauss. But within any closed model universe with the topology of a 3-sphere, a Gaussian 2-sphere that is expanded widely enough from one point finds itself collapsing to nothingness at the antipodal point. Also collapsed to nothingness is the attempt to acquire useful information about the "charge of the universe": the charge is trivially zero. By the same token, every "surface integral" (see details in Chapter 20) to determine mass-energy or angular momentum collapses to nothingness. To make the same point in another way: around a closed universe there is no place to put a test object or gyroscope into Keplerian orbit to determine either any so-called "total mass" or "rest frame" or "4-momentum" or "angular momentum" of the system. These terms are undefined and undefinable. Words, yes; meaning, no.

Not having a defined 4-momentum for a closed universe may seem at first sight disturbing; but it would be far more disturbing to be given four numbers and to be told authoritatively that they represent the components of some purported "total energy-momentum 4-vector of the universe." Components with respect to what local Lorentz frame? At what point? And what about the change in this vector on parallel transport around a closed path leading back to that strangely preferred point? It is a happy salvation from these embarrassments that the issue does not and cannot arise!

Imagine a fantastically precise measurement of the energy of a γ-ray. The experimenter wishes to know how much this γ-ray contributes to the total mass-energy of the universe. Having measured its energy in the laboratory, he then corrects it for the negative gravitational energy by which it is bound to the Earth. The result,

$$E_{\text{corrected}} = h\nu(1 - M_{\oplus}/R_{\oplus}),$$

is the energy the photon will have after it climbs out of the Earth's gravitational field. But this is only the first in a long chain of corrections for energy losses (redshifts) as the photon climbs out of the gravitational fields of the solar system, the galaxy, the local cluster of galaxies, the supercluster, and then what? These corrections show no sign of converging, unless to $E_{\text{corrected}} = 0$.

Asymptotic flatness as the key to the definability of M and S

Quite in contrast to the charge-energy-angular-momentum facelessness of a closed universe are the attractive possibilities of defining and measuring all three quantities in any space that is asymptotically flat. One does not have to revolutionize present-day views of cosmology to talk of asymptotically flat space. It is enough to note how small is the departure from flatness, as measured by the departure of $(-g_{00})^{1/2}$ from unity, in cases of astronomical or astrophysical interest (Box 19.2). Surrounding a region where any dynamics, however complicated, is going on, whenever the geometry is asymptotically flat to some specified degree of precision, then to that degree of precision it makes sense to speak of the total energy-momentum 4-vector of the dynamic region, $\boldsymbol{P}$, and its total intrinsic angular momentum, $\boldsymbol{S}$. Parallel transport of either around any closed curve in the flat region brings it back to its

Box 19.2 METRIC CORRECTION TERM NEAR SELECTED HEAVENLY BODIES

	m		m	r	$\frac{m}{r} = 1 - (-g_{00})^{1/2}$
At shoulder of Venus de Milo	2×10^5 g	=	1.5×10^{-23} cm	30 cm	5×10^{-25}
At surface of Earth	6×10^{27} g	=	4×10^{-1} cm	6.4×10^8 cm	6×10^{-10}
At Earth's distance from sun	2×10^{33} g	=	1.5×10^5 cm	1.5×10^{13} cm	1×10^{-8}
At sun's distance from center of galaxy	2×10^{44} g	=	1.5×10^{16} cm	2.5×10^{22} cm	6×10^{-7}
At distance of galaxy from center of Virgo cluster of galaxies	6×10^{47} g	=	4×10^{19} cm	3×10^{25} cm	1×10^{-6}

starting point unchanged. Moreover, it makes no difference how enormous are the departures from flatness in the dynamic region (black holes, collapsing stars, intense gravitational waves, etc.); far away the curvature will be weak, and the 4-momentum and angular momentum will reveal themselves by their imprints on the spacetime geometry.

2

CHAPTER 20

CONSERVATION LAWS FOR 4-MOMENTUM AND ANGULAR MOMENTUM

We denote as energy of a material system in a certain state the contribution of all effects (measured in mechanical units of work) produced outside the system when it passes in an arbitrary manner from its state to a reference state which has been defined ad hoc.

WILLIAM THOMPSON (later Lord Kelvin),
as quoted by Max von Laue in Schilpp (1949), p. 514.

All *forms of energy possess inertia.*

ALBERT EINSTEIN, conclusion
from his paper of September 26, 1905,
as summarized by von Laue in Schilpp (1949), p. 523.

§20.1. OVERVIEW

Chapter 5 (stress-energy tensor) is needed as preparation for this chapter, which in turn is needed as preparation for the Track-2 portion of Chapter 36 (generation of gravitational waves) and will be useful in understanding Chapter 35 (propagation of gravitational waves).

Chapter 19 expounded the key features of total 4-momentum $\boldsymbol{P}$ and total angular momentum $\boldsymbol{S}$ for an arbitrary, gravitating system. But one crucial feature was left unproved: the conservation laws for $\boldsymbol{P}$ and $\boldsymbol{S}$ (Box 19.1). To prove those conservation laws is the chief purpose of this chapter. But other interesting, if less important, aspects of $\boldsymbol{P}$ and $\boldsymbol{S}$ will be encountered along the route to the proof—Gaussian flux integrals for 4-momentum and angular momentum; a stress-energy "pseudotensor" for the gravitational field, which is a tool in constructing volume integrals for $\boldsymbol{P}$ and $\boldsymbol{S}$; and the nonlocalizability of the energy of the gravitational field.

§20.2. GAUSSIAN FLUX INTEGRALS FOR 4-MOMENTUM AND ANGULAR MOMENTUM

Gaussian flux integrals for charge and Newtonian mass

In electromagnetic theory one can determine the conserved total charge of a source by adding up the number of electric field lines emanating from it—i.e., by performing a Gaussian flux integral over a closed 2-surface surrounding it:

$$Q = \frac{1}{4\pi}\oint E^j \, d^2S_j = \frac{1}{4\pi}\oint F^{0j}\, d^2S_j. \tag{20.1}$$

Similarly, in Newtonian theory one can determine the mass of a source by evaluating the Gaussian flux integral

$$M = \frac{1}{4\pi}\oint \Phi_{,j}\, d^2S_j. \tag{20.2}$$

These flux integrals work because the charge and mass of a source place indelible imprints on the electromagnetic and gravitational fields that envelop it.

The external gravitational field (spacetime geometry) in general relativity possesses similar imprints, imprints not only of the source's total mass-energy M, but also of its total 4-momentum $\boldsymbol{P}$ and its intrinsic angular momentum $\boldsymbol{S}$ (see Box 19.1). Hence, it is reasonable to search for Gaussian flux integrals that represent the 4-momentum and angular momentum of the source.

To simplify the search, carry it out initially in linearized theory, and use Maxwell electrodynamics as a guide. In electrodynamics the Gaussian flux integral for charge follows from Maxwell's equations $F^{\mu\nu}{}_{,\nu} = 4\pi J^\mu$, plus the crucial fact that $F^{\mu\nu}$ is antisymmetric, so that $F^{0\mu}{}_{,\mu} = F^{0j}{}_{,j}$:

$$Q = \int J^0\, d^3x = \frac{1}{4\pi}\int F^{0\nu}{}_{,\nu}\, d^3x = \frac{1}{4\pi}\int F^{0j}{}_{,j}\, d^3x \underset{\text{[Gauss's theorem]}}{=} \frac{1}{4\pi}\oint F^{0j}\, d^2S_j.$$

To find analogous flux integrals in linearized theory, rewrite the linearized field equations (18.7) in an analogous form involving an entity with analogous crucial symmetries. The entity needed turns out to be

$H^{\mu\alpha\nu\beta}$ defined

$$H^{\mu\alpha\nu\beta} \equiv -(\bar{h}^{\mu\nu}\eta^{\alpha\beta} + \eta^{\mu\nu}\bar{h}^{\alpha\beta} - \bar{h}^{\alpha\nu}\eta^{\mu\beta} - \bar{h}^{\mu\beta}\eta^{\alpha\nu}). \tag{20.3}$$

As one readily verifies from this expression, it has the same symmetries as the Riemann tensor

$$\begin{aligned} H^{\mu\alpha\nu\beta} &= H^{\nu\beta\mu\alpha} = H^{[\mu\alpha][\nu\beta]}, \\ H^{\mu[\alpha\nu\beta]} &= 0. \end{aligned} \tag{20.4}$$

This entity, like $\bar{h}^{\mu\nu}$, transforms as a tensor under the Lorentz transformations of linearized theory; but it is not gauge-invariant, so it is not a tensor in the general relativistic sense.

2

Linearized field equations in terms of $H^{\mu\alpha\nu\beta}$

In terms of $H^{\mu\alpha\nu\beta}$, the linearized field equations (18.7) take on the much simplified form

$$2G^{\mu\nu} = H^{\mu\alpha\nu\beta}{}_{,\alpha\beta} = 16\pi T^{\mu\nu}; \tag{20.5}$$

and from these, by antisymmetry of $H^{\mu\alpha\nu\beta}$ in ν and β, follow the source conservation laws of linearized theory,

$$T^{\mu\nu}{}_{,\nu} = \frac{1}{16\pi} H^{\mu\alpha\nu\beta}{}_{,\alpha\beta\nu} = 0,$$

which were discussed back in §18.3. The same antisymmetry as yields these equations of motion also produces a Gaussian flux integral for the source's total 4-momentum:

Gaussian flux integrals in linearized theory: (1) for 4-momentum

$$\begin{aligned} P^\mu &= \int T^{\mu 0}\, d^3x = \frac{1}{16\pi}\int H^{\mu\alpha 0\beta}{}_{,\alpha\beta}\, d^3x = \frac{1}{16\pi}\int H^{\mu\alpha 0j}{}_{,\alpha j}\, d^3x \\ &= \frac{1}{16\pi}\oint_{\mathcal{S}} H^{\mu\alpha 0j}{}_{,\alpha}\, d^2S_j. \\ &\quad \text{[Gauss's theorem]} \end{aligned} \tag{20.6}$$

Here the closed 2-surface of integration $\mathcal{S}$ must completely surround the source and must lie in a 3-surface of constant time x^0. The integral (20.6) for the source's energy P^0, which is used more frequently than the integrals for P^j, reduces to an especially simple form in terms of $g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}$:

$$P^0 = \frac{1}{16\pi}\int_{\mathcal{S}} (g_{jk,k} - g_{kk,j})\, d^2S_j \tag{20.7}$$

(see exercise 20.1).

A calculation similar to (20.6), but more lengthy (exercise 20.2), yields a flux integral for total angular momentum about the origin of coordinates:

(2) for angular momentum

$$\begin{aligned} J^{\mu\nu} &= \int (x^\mu T^{\nu 0} - x^\nu T^{\mu 0})\, d^3x \\ &= \frac{1}{16\pi}\oint_{\mathcal{S}} (x^\mu H^{\nu\alpha 0j}{}_{,\alpha} - x^\nu H^{\mu\alpha 0j}{}_{,\alpha} + H^{\mu j 0\nu} - H^{\nu j 0\mu})\, d^2S_j. \end{aligned} \tag{20.8}$$

To evaluate the flux integrals in (20.6) to (20.8) (by contrast with the volume integrals), one need utilize only the gravitational field far outside the source. Since that gravitational field has the same form in full general relativity for strong sources as in linearized theory for weak sources, the flux integrals can be used to calculate P^μ and $J^{\mu\nu}$ for any isolated source whatsoever, weak or strong:

Generalization of Gaussian flux integrals to full general relativity

$$\left.\begin{aligned} P^\mu &= \frac{1}{16\pi}\oint_{\mathcal{S}} H^{\mu\alpha 0j}{}_{,\alpha}\, d^2S_j, \\ P^0 &= \frac{1}{16\pi}\oint_{\mathcal{S}} (g_{jk,k} - g_{kk,j})\, d^2S_j, \\ J^{\mu\nu} &= \frac{1}{16\pi}\oint_{\mathcal{S}} (x^\mu H^{\nu\alpha 0j}{}_{,\alpha} - x^\nu H^{\mu\alpha 0j}{}_{,\alpha} \\ &\qquad + H^{\mu j 0\nu} - H^{\nu j 0\mu})\, d^2S_j. \end{aligned}\right\} \begin{array}{l}\text{in full general relativity}\\ \text{theory, for any isolated}\\ \text{source, when the closed}\\ \text{surface of integration } \mathcal{S}\\ \text{is in the asymptotically}\\ \text{flat region surrounding}\\ \text{the source, and when}\\ \text{asymptotically Minkows-}\\ \text{kian coordinates are used.}\end{array} \tag{20.9}$$

Knowing P^μ and $J^{\mu\nu}$, one can calculate the source's total mass-energy M and intrinsic angular momentum S^μ by the standard procedure of Box 5.6:

Total mass-energy, center of mass, and intrinsic angular momentum

$$M = (-P^\mu P_\mu)^{1/2}, \tag{20.10}$$

$$Y^\mu = -J^{\mu\nu}P_\nu/M^2 = \begin{pmatrix}\text{vector by which the source's asymptotic,}\\ \text{"}M/r\text{", spherical field is displaced from}\\ \text{being centered on the origin of coordinates}\end{pmatrix} \tag{20.11}$$

$$S_\rho = \frac{1}{2}\epsilon_{\mu\nu\sigma\rho}(J^{\mu\nu} - Y^\mu P^\nu + Y^\nu P^\mu)P^\sigma/M. \tag{20.12}$$

Note especially that *the integrands of the flux integrals (20.9) are not gauge-invariant*. In any local inertial frame at an event $\mathscr{P}_0$ $[g_{\mu\nu}(\mathscr{P}_0) = \eta_{\mu\nu}, g_{\mu\nu,\alpha}(\mathscr{P}_0) = 0]$ they vanish, since

$$g_{\mu\nu,\alpha} = h_{\mu\nu,\alpha} = 0 \Rightarrow H^{\mu\nu\alpha\beta}{}_{,\alpha} = 0; \qquad g_{\mu\nu} = \eta_{\mu\nu} \Rightarrow H^{\mu\nu\alpha\beta} = 0.$$

This is reasonable behavior; their Newtonian analog, the integrand $\Phi_{,j}$ = (gravitational acceleration) of the Newtonian flux integral (20.2), similarly vanishes in local inertial frames.

Although the integrands of the flux integrals are not gauge-invariant, the total integrals P^μ (4-momentum) and $J^{\mu\nu}$ (angular momentum) most assuredly are! They have meaning and significance independent of any coordinate system and gauge. They are tensors in the asymptotically flat region surrounding the source.

The spacetime must be asymptotically flat if there is to be any possibility of defining energy and angular momentum. Only then can linearized theory be applied; and only on the principle that linearized theory applies far away can one justify using the flux integrals (20.9) in the full nonlinear theory. Nobody can compel a physicist to move in close to define energy and angular momentum. He has no need to move in close; and he may have compelling motives not to: the internal structure of the sources may be inaccessible, incomprehensible, uninteresting, dangerous, expensively distant, or frightening. This requirement for far-away flatness is a remarkable feature of the flux integrals (20.9); it is also a decisive feature. Even the coordinates must be asymptotically Minkowskian; otherwise most formulas in this chapter fail or require modification. In particular, *when evaluating the 4-momentum and angular momentum of a localized system, one must apply the flux integrals (20.9) only in asymptotically Minkowskian coordinates. If such coordinates do not exist (spacetime not flat at infinity), one must completely abandon the flux integrals, and the quantities that rely on them for definition: the total mass, momentum, and angular momentum of the gravitating source.* In this connection, recall the discussion of §19.4. It described, in physical terms, why "total mass-energy" is a limited concept, useful only when one adopts a limited viewpoint that ignores cosmology. (Compare "light ray" or "particle," concepts of enormous value, but concepts that break down when wave optics or wave mechanics enter significantly.)

Gaussian flux integrals valid only in asymptotically flat region of spacetime and in asymptotically Minkowskian coordinates

Summary: Attempts to use formulas (20.9) in ways that lose sight of the Minkowski boundary conditions (and especially simply adopting them unmodified in curvilinear coordinates) easily and unavoidably produce nonsense.

EXERCISES

Exercise 20.1. FLUX INTEGRAL FOR TOTAL MASS-ENERGY IN LINEARIZED THEORY

Show that the flux integral (20.6) for P^0 reduces to (20.7). Then show that, when applied to a nearly Newtonian source [line element (18.15c)], it reduces further to the familiar Newtonian flux integral (20.2).

Exercise 20.2. FLUX INTEGRAL FOR ANGULAR MOMENTUM IN LINEARIZED THEORY

Derive the Gaussian flux integral (20.8) for $J^{\mu\nu}$. [*Hint:* use the field equations (20.5) to show

$$16\pi x^\mu T^{\nu 0} = (x^\mu H^{\nu\alpha 0k}{}_{,\alpha})_{,k} - H^{\nu j 0\mu}{}_{,j} - H^{\nu 00\mu}{}_{,0}; \tag{20.13}$$

and then use Gauss's theorem to evaluate the volume integral of equation (20.8)].

Exercise 20.3. FLUX INTEGRALS FOR AN ARBITRARY STATIONARY SOURCE

(a) Use the flux integrals (20.9) to calculate P^μ and $J^{\mu\nu}$ for an arbitrary stationary source. For the asymptotically flat metric around the source, use (19.13), with the gravitational radiation terms set to zero.

(b) Verify that the "auxiliary equations" (20.10) to (20.12) give the correct answer for this source's total mass-energy M and intrinsic angular momentum S^μ.

§20.3. VOLUME INTEGRALS FOR 4-MOMENTUM AND ANGULAR MOMENTUM

It is easy, in linearized theory, to convert the surface integrals for P^μ and $J^{\mu\nu}$ into volume integrals over the source; one can simply trace backward the steps that led to the surface integrals in the first place [equation (20.6); exercise 20.2]. How, in full general relativity, can one similarly convert from the surface integrals to volume integrals? The answer is rather easy, if one thinks in the right direction. One need only put the full Einstein field equations into the form

The full Einstein field equations in terms of $H^{\mu\alpha\nu\beta}$

$$H^{\mu\alpha\nu\beta}{}_{,\alpha\beta} = 16\pi T^{\mu\nu}_{\text{eff}} \tag{20.14}$$

analogous to equations (20.5) of linearized theory. Here $H^{\mu\alpha\nu\beta}$ is to be defined in terms of $h_{\mu\nu} \equiv g_{\mu\nu} - \eta_{\mu\nu}$ by equation (20.3), even deep inside the source where $|h_{\mu\nu}|$ might be $\gtrsim 1$. This form of the Einstein equations then permits a conversion of the Gaussian flux integrals into volume integrals, just as in linearized theory:

Volume integrals for 4-momentum and angular momentum in full general relativity

$$\begin{aligned} P^\mu &= \frac{1}{16\pi}\oint H^{\mu\alpha 0j}{}_{,\alpha}\, d^2S_j = \frac{1}{16\pi}\int H^{\mu\alpha 0j}{}_{,\alpha j}\, d^3x = \frac{1}{16\pi}\int H^{\mu\alpha 0\beta}{}_{,\alpha\beta}\, d^3x \\ &= \int T^{\mu 0}_{\text{eff}}\, d^3x. \end{aligned} \tag{20.15}$$

Similarly,

$$J^{\mu\nu} = \int (x^\mu T^{\nu 0}_{\text{eff}} - x^\nu T^{\mu 0}_{\text{eff}})\, d^3x. \tag{20.16}$$

[Crucial to the conversion is the use of partial derivatives rather than covariant derivatives in equations (20.14).] In these volume integrals, as throughout the preceeding discussion, the coordinates must become asymptotically Lorentz ($g_{\mu\nu} \to \eta_{\mu\nu}$) far from the source.

The form of $T^{\mu\nu}_{\text{eff}}$ can be calculated by recalling that $H^{\mu\alpha\nu\beta}{}_{,\alpha\beta}$ is a linearized approximation to the Einstein curvature tensor (20.5). Define the nonlinear corrections by

$$16\pi t^{\mu\nu} \equiv H^{\mu\alpha\nu\beta}{}_{,\alpha\beta} - 2G^{\mu\nu}. \tag{20.17}$$

$t^{\mu\nu}$ ("stress-energy pseudotensor") defined

(To calculate them in terms of $g_{\mu\nu}$ or $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$ is straightforward but lengthy. The precise form of these corrections will never be needed in this book.) Then Einstein's equations read

$$H^{\mu\alpha\nu\beta}{}_{,\alpha\beta} = 16\pi t^{\mu\nu} + 2G^{\mu\nu} = 16\pi(t^{\mu\nu} + T^{\mu\nu}),$$

so that

$$T^{\mu\nu}_{\text{eff}} = T^{\mu\nu} + t^{\mu\nu}. \tag{20.18}$$

$T^{\mu\nu}_{\text{eff}}$ defined

The quantity $t^{\mu\nu}$ is sometimes called a "stress-energy pseudotensor for the gravitational field." The Einstein field equations (20.14) imply, because $H^{\mu\alpha\nu\beta}{}_{,\alpha\beta}$ is antisymmetric in ν and β, that

$$T^{\mu\nu}_{\text{eff},\nu} = (T^{\mu\nu} + t^{\mu\nu})_{,\nu} = 0. \tag{20.19}$$

Conservation law for $T^{\mu\nu}_{\text{eff}}$

These equations are equivalent to $T^{\mu\nu}{}_{;\nu} = 0$, but they are written with partial derivatives rather than covariant derivatives—a fact that permits conversions back and forth between volume integrals and surface integrals.

$H^{\mu\alpha\nu\beta}$, $t^{\mu\nu}$, and $T^{\mu\nu}_{\text{eff}}$ are coordinate-dependent objects

All the quantities $H^{\mu\alpha\nu\beta}$, $T^{\mu\nu}_{\text{eff}}$, and $t^{\mu\nu}$ depend for their definition and existence on the choice of coordinates; they have no existence independent of coordinates; they are not components of tensors or of any other geometric object. Correspondingly, the equations (20.14) to (20.19) involving $T^{\mu\nu}_{\text{eff}}$ and $t^{\mu\nu}$ have no geometric, coordinate-free significance; they are not "covariant tensor equations." There is, nevertheless, adequate invariance under general coordinate transformations to give the values P^{μ} and $J^{\mu\nu}$ of the volume integrals (20.15) and (20.16) geometric, coordinate-free significance in the asymptotically flat region far outside the source. Although this invariance is hard to see in the volume integrals themselves, it is clear from the surface-integral forms (20.9) that no coordinate transformation which changes the coordinates only inside some spatially bounded region can influence the values of the integrals. For coordinate changes in the distant, asymptotically flat regions, linearized theory guarantees that under Lorentz transformations the integrals for P^{μ} and $J^{\mu\nu}$ will transform like special relativistic tensors, and that under infinitesimal coordinate transformations (gauge changes) they will be invariant.

Because $t^{\mu\nu}$ are not tensor components, they can vanish at a point in one coordinate system but not in another. The resultant ambiguity in defining a localized energy density t^{00} for the gravitational field has a counterpart in ambiguities that exist in

Other, equally good versions of $H^{\mu\alpha\nu\beta}$, $t^{\mu\nu}$, $T^{\mu\nu}_{\text{eff}}$:

the formal definition of $t^{\mu\nu}$. It is clear that any quantities $H^{\mu\alpha\nu\beta}_{\text{new}}$ which agree with the original $H^{\mu\alpha\nu\beta}$ in the asymptotic weak-field region will give the same values as $H^{\mu\alpha\nu\beta}$ does for the P^{μ} and $J^{\mu\nu}$ surface integrals (20.9). One especially convenient choice has been given by Landau and Lifshitz (1962; §100), who define

(1) $H^{\mu\alpha\nu\beta}_{\text{L-L}}$

$$H^{\mu\alpha\nu\beta}_{\text{L-L}} = \mathfrak{g}^{\mu\nu}\mathfrak{g}^{\alpha\beta} - \mathfrak{g}^{\alpha\nu}\mathfrak{g}^{\mu\beta}, \tag{20.20}$$

where $\mathfrak{g}^{\mu\nu} \equiv (-g)^{1/2} g^{\mu\nu}$. Landau and Lifshitz show that Einstein's equations can be written in the form

$$H^{\mu\alpha\nu\beta}_{\text{L-L},\alpha\beta} = 16\pi(-g)(T^{\mu\nu} + t^{\mu\nu}_{\text{L-L}}), \tag{20.21}$$

(2) $t^{\alpha\beta}_{\text{L-L}}$

where the Landau-Lifshitz pseudotensor components

$$\begin{aligned}(-g)\, t^{\alpha\beta}_{\text{L-L}} = \frac{1}{16\pi}\Big\{ & \mathfrak{g}^{\alpha\beta}{}_{,\lambda}\mathfrak{g}^{\lambda\mu}{}_{,\mu} - \mathfrak{g}^{\alpha\lambda}{}_{,\lambda}\mathfrak{g}^{\beta\mu}{}_{,\mu} + \frac{1}{2} g^{\alpha\beta} g_{\lambda\mu}\mathfrak{g}^{\lambda\nu}{}_{,\rho}\mathfrak{g}^{\rho\mu}{}_{,\nu} \\ & - (g^{\alpha\lambda}g_{\mu\nu}\mathfrak{g}^{\beta\nu}{}_{,\rho}\mathfrak{g}^{\mu\rho}{}_{,\lambda} + g^{\beta\lambda}g_{\mu\nu}\mathfrak{g}^{\alpha\nu}{}_{,\rho}\mathfrak{g}^{\mu\rho}{}_{,\lambda}) + g_{\lambda\mu}g^{\nu\rho}\mathfrak{g}^{\alpha\lambda}{}_{,\nu}\mathfrak{g}^{\beta\mu}{}_{,\rho} \\ & + \frac{1}{8}(2g^{\alpha\lambda}g^{\beta\mu} - g^{\alpha\beta}g^{\lambda\mu})(2g_{\nu\rho}g_{\sigma\tau} - g_{\rho\sigma}g_{\nu\tau})\mathfrak{g}^{\nu\tau}{}_{,\lambda}\mathfrak{g}^{\rho\sigma}{}_{,\mu}\Big\}\end{aligned} \tag{20.22}$$

are precisely quadratic in the first derivatives of the metric. (Einstein also gave a pseudotensor $t_{E}{}^{\mu}{}_{\nu}$ with this property, but it was not symmetric and so did not lead to an integral for $J^{\mu\nu}$.) Because $H^{\mu\alpha\nu\beta}_{\text{L-L}}$ has the same symmetries as $H^{\mu\alpha\nu\beta}$ and equals $H^{\mu\alpha\nu\beta}$ far from the source (exercise 20.4), and because the field equations (20.21) in terms of $H^{\mu\alpha\nu\beta}_{\text{L-L}}$ have the same form as in terms of $H^{\mu\alpha\nu\beta}$, it follows that

(3) $T^{\mu\nu}_{\text{L-L eff}}$

$$T^{\mu\nu}_{\text{L-L eff}} \equiv (-g)(T^{\mu\nu} + t^{\mu\nu}_{\text{L-L}}) \tag{20.23a}$$

has all the properties of the $T^{\mu\nu}_{\text{eff}}$ introduced earlier in this section:

$$T^{\mu\nu}_{\text{L-L eff},\nu} = 0, \tag{20.23b}$$

$$P^{\mu} = \int T^{\mu 0}_{\text{L-L eff}}\, d^3x, \tag{20.23c}$$

$$J^{\mu\nu} = \int (x^{\mu} T^{\nu 0}_{\text{L-L eff}} - x^{\nu} T^{\mu 0}_{\text{L-L eff}})\, d^3x. \tag{20.23d}$$

EXERCISE

Exercise 20.4. FORM OF $H^{\mu\alpha\nu\beta}_{\text{L-L}}$ FAR FROM SOURCE

Show that the entities $H^{\mu\alpha\nu\beta}_{\text{L-L}}$ of equations (20.20) reduce to $H^{\mu\alpha\nu\beta}$ (20.3) in the weak-field region far outside the source.

---

§20.4. WHY THE ENERGY OF THE GRAVITATIONAL FIELD CANNOT BE LOCALIZED

Consider an element of 3-volume $d\Sigma_{\nu}$ and evaluate the contribution of the "gravitational field" in that element of 3-volume to the energy-momentum 4-vector, using in the calculation either the pseudotensor $t^{\mu\nu}$ or the pseudotensor $t^{\mu\nu}_{\text{L-L}}$ discussed in the last section. Thereby obtain

$$\boldsymbol{p} = \boldsymbol{e}_\mu t^{\mu\nu}\, d\Sigma_\nu$$

or

$$\boldsymbol{p} = \boldsymbol{e}_\mu t^{\mu\nu}_{\text{L-L}}\, d\Sigma_\nu.$$

Right? No, the question is wrong. The motivation is wrong. The result is wrong. The idea is wrong.

To ask for the amount of electromagnetic energy and momentum in an element of 3-volume makes sense. First, there is one and only one formula for this quantity. Second, and more important, this energy-momentum in principle "has weight." It curves space. It serves as a source term on the righthand side of Einstein's field equations. It produces a relative geodesic deviation of two nearby world lines that pass through the region of space in question. It is observable. Not one of these properties does "local gravitational energy-momentum" possess. There is no unique formula for it, but a multitude of quite distinct formulas. The two cited are only two among an infinity. Moreover, "local gravitational energy-momentum" has no weight. It does not curve space. It does not serve as a source term on the righthand side of Einstein's field equations. It does not produce any relative geodesic deviation of two nearby world lines that pass through the region of space in question. It is not observable.

Why one cannot define a localized energy-momentum for the gravitational field

Anybody who looks for a magic formula for "local gravitational energy-momentum" is looking for the right answer to the wrong question. Unhappily, enormous time and effort were devoted in the past to trying to "answer this question" before investigators realized the futility of the enterprise. Toward the end, above all mathematical arguments, one came to appreciate the quiet but rock-like strength of Einstein's equivalence principle. One can always find in any given locality a frame of reference in which all local "gravitational fields" (all Christoffel symbols; all $\Gamma^\alpha{}_{\mu\nu}$) disappear. No Γ's means no "gravitational field" and no local gravitational field means no "local gravitational energy-momentum."

Nobody can deny or wants to deny that gravitational forces make a contribution to the mass-energy of a gravitationally interacting system. The mass-energy of the Earth-moon system is less than the mass-energy that the system would have if the two objects were at infinite separation. The mass-energy of a neutron star is less than the mass-energy of the same number of baryons at infinite separation. Surrounding a region of empty space where there is a concentration of gravitational waves, there is a net attraction, betokening a positive net mass-energy in that region of space (see Chapter 35). At issue is not the existence of gravitational energy, but the localizability of gravitational energy. It is not localizable. The equivalence principle forbids.

Look at an old-fashioned potato, replete with warts and bumps. With an orange marking pen, mark on it a "North Pole" and an "equator". The length of the equator is very far from being equal to 2π times the distance from the North Pole to the

2

equator. The explanation, "curvature," is simple, just as the explanation, "gravitation", for the deficit in mass of the earth-moon system (or deficit for the neutron star, or surplus for the region of space occupied by the gravitational waves) is simple. Yet it is not possible to ascribe the deficit in the length of the equator in the one case, or in mass in the other case, in any uniquely right way to different elements of the manifold (2-dimensional in the one case, 3-dimensional in the other). Look at a small region on the surface of the potato. The geometry there is locally flat. Look at any small region of space in any of the three gravitating systems. In an appropriate coordinate system it is free of gravitational field. The over-all effect one is looking at is a global effect, not a local effect. That is what the mathematics cries out. That is the lesson of the nonuniqueness of the $t^{\mu\nu}$!

§20.5. CONSERVATION LAWS FOR TOTAL 4-MOMENTUM AND ANGULAR MOMENTUM

Consider a system such as our galaxy or the solar system, which is made up of many gravitating bodies. Some of the bodies may be highly relativistic (black holes; neutron stars), while others are not. However, insist that in the regions between the bodies spacetime be nearly flat (gravity be weak)—so flat, in fact, that one can cover the entire system with coordinates which are (almost) globally inertial, except in a small neighborhood of each body where gravity may be strong. Such coordinates can exist only if the Newtonian gravitational potential, $\Phi \approx \frac{1}{2}(\eta_{00} - g_{00})$, in the interbody region is small:

$$\Phi_{\text{interbody}} \sim (\text{Mass of system})/(\text{radius of system}) \ll 1.$$

The solar system certainly satisfies this condition ($\Phi_{\text{interbody}} \sim 10^{-7}$), as does the Galaxy ($\Phi_{\text{interbody}} \sim 10^{-6}$), as do clusters of galaxies ($\Phi_{\text{interbody}} \sim 10^{-6}$); *but the universe as a whole does not* ($\Phi_{\text{interbody}} \sim 1$)!

In evaluating volume integrals for the system's total 4-momentum, split its volume into a region containing each body (denoted "A") plus an interbody region; and neglect the pseudotensor contribution from the almost-flat interbody region:

Total 4-momentum and angular momentum for a system of gravitating bodies

$$\begin{aligned} P^{\mu}_{\text{system}} &= \sum_A \int_A T^{\mu 0}_{\text{eff}}\, d^3x + \int_{\substack{\text{interbody}\\ \text{region}}} T^{\mu 0}_{\text{eff}}\, d^3x \\ &= \sum_A P^{\mu}_A + \int_{\substack{\text{interbody}\\ \text{region}}} T^{\mu 0}\, d^3x. \end{aligned} \tag{20.24a}$$

Because spacetime is asymptotically flat around each body, $P_A{}^{\mu}$ is the 4-momentum of body A as measured gravitationally by an experimenter near it. The integral of $T^{\mu 0}$ over the interbody region is the contribution of any gas, particles, or magnetic fields out there to the total 4-momentum. A similar breakup of the angular momentum reads

$$J^{\mu\nu}_{\text{system}} = \sum_A J^{\mu\nu}_A + \int_{\substack{\text{interbody}\\ \text{region}}} (x^\mu T^{\nu 0} - x^\nu T^{\mu 0})\, d^3x. \tag{20.24b}$$

In operational terms, these breakups show that *the total 4-momentum and angular momentum of the system, as measured gravitationally by an experimenter outside it, are sums of P^μ and $J^{\mu\nu}$ for each individual body, as measured gravitationally by an experimenter near it, plus contributions of the usual special-relativistic type from the interbody matter and fields.* This is true even if some of the bodies are hurtling through the system with speeds near that of light; their gravitationally measured P^μ and $J^{\mu\nu}$ contribute, on an equal footing with anyone else's, to the system's total P^μ and $J^{\mu\nu}$!

Rates of change of total 4-momentum and angular momentum:

Surround this asymptotically flat system by a two-dimensional surface $\mathcal{S}$ that is at rest in some asymptotic Lorentz frame. Then the 4-momentum and angular momentum inside $\mathcal{S}$ change at a rate (as measured in $\mathcal{S}$'s rest frame) given by

(1) expressed as flux integrals of $T^{\mu\nu}_{\text{eff}}$

$$\begin{aligned}\frac{dP^\mu}{dt} &= \frac{d}{dt}\int T^{\mu 0}_{\text{eff}}\, d^3x = \int T^{\mu 0}_{\text{eff},0}\, d^3x = -\int T^{\mu j}_{\text{eff},j}\, d^3x \\ &= -\oint T^{\mu j}_{\text{eff}}\, d^2S_j,\end{aligned} \tag{20.25}$$

and similarly

$$\frac{dJ^{\mu\nu}}{dt} = -\oint_{\mathcal{S}_2} (x^\mu T^{\nu j}_{\text{eff}} - x^\nu T^{\mu j}_{\text{eff}})\, d^2S_j. \tag{20.26}$$

Although the pseudotensor $t^{\mu\nu}$, in the interbody region and outside the system, contributes negligibly to the total 4-momentum and angular momentum (by assumption), its contribution via gravitational waves to the time derivatives dP^μ/dt and $dJ^{\mu\nu}/dt$ can be important when added up over astronomical periods of time. Thus, one must not ignore it in the flux integrals (20.25), (20.26).

In evaluating these flux integrals, it is especially convenient to use the Landau-Lifshitz form of $T^{\mu\nu}_{\text{eff}}$, since that form contains no second derivatives of the metric. Thus set

$$T^{\mu\nu}_{\text{eff}} = (-g)(T^{\mu\nu} + t^{\mu\nu}_{\text{L-L}}) \approx (T^{\mu\nu} + t^{\mu\nu}_{\text{L-L}}),$$

where $t^{\mu\nu}_{\text{L-L}}$ are given by equations (20.22). Only those portions of $t^{\mu\nu}_{\text{L-L}}$ that die out as $1/r^2$ or $1/r^3$ at large r can contribute to the flux integrals (20.25), (20.26). For static solutions $[\mathfrak{g}_{\mu\nu} \sim \text{const.} + O(1/r)]$, $t^{\mu\nu}_{\text{L-L}}$ dies out as $1/r^4$. Hence, the only contributions come from dynamic parts of the metric, which, at these large distances, are entirely in the form of gravitational waves. The study of gravitational waves in Chapter 35 will reveal that when $t^{\mu\nu}_{\text{L-L}}$ is averaged over several wavelengths, it becomes a stress-energy tensor $T^{(\text{GW})\mu\nu}$ for the waves, which has all the properties one ever requires of any stress-energy tensor. (For example, via Einstein's equations

$G^{(B)\mu\nu} = 8\pi T^{(GW)\mu\nu}$, it contributes to the "background" curvature of the spacetime through which the waves propagate.) Moreover, averaging $t^{\mu\nu}_{\text{L-L}}$ over several wavelengths before evaluating the flux integrals (20.25), (20.26) cannot affect the values of the integrals. Therefore, one can freely make in these integrals the replacement

$$T^{\mu\nu}_{\text{eff}} = T^{\mu\nu} + T^{(GW)\mu\nu},$$

thereby obtaining

(2) expressed as flux integrals of $T^{\mu\nu} + T^{(GW)\mu\nu}$

$$-\frac{dP^{\mu}}{dt} = \oint_S (T^{\mu j} + T^{(GW)\mu j})\, d^2S_j, \tag{20.27}$$

$$-\frac{dJ^{\mu\nu}}{dt} = \oint_S [x^{\mu}(T^{\nu j} + T^{(GW)\nu j}) - x^{\nu}(T^{\mu j} + T^{(GW)\mu j})]\, d^2S_j. \tag{20.28}$$

These are tensor equations in the asymptotically flat spacetime surrounding the system. All reference to pseudotensors and other nontensorial entities has disappeared.

Equations (20.27) and (20.28) say that *the rate of loss of 4-momentum and angular momentum from the system, as measured gravitationally, is precisely equal to the rate at which matter, fields, and gravitational waves carry off 4-momentum and angular momentum.*

This theorem is extremely useful in thought experiments where one imagines changing the 4-momentum or angular momentum of a highly relativistic body (e.g., a rotating neutron star) by throwing particles onto it from far away [see, e.g., Hartle (1970)].

EXERCISE

Exercise 20.5. TOTAL MASS-ENERGY IN NEWTONIAN LIMIT

(a) Calculate $t^{\alpha\beta}_{\text{L-L}}$ for the nearly Newtonian metric

$$ds^2 = -(1 + 2\Phi)\, dt^2 + (1 - 2\Phi)\, \delta_{jk}\, dx^j\, dx^k$$

(see §18.4). Assume the source is slowly changing, so that time derivatives of Φ can be neglected compared to space derivatives. [*Answer:*

$$t^{00}_{\text{L-L}} = -\frac{7}{8\pi}\Phi_{,j}\Phi_{,j},$$

$$t^{0j}_{\text{L-L}} = 0,$$

$$t^{jk}_{\text{L-L}} = \frac{1}{4\pi}(\Phi_{,j}\Phi_{,k} - \frac{1}{2}\,\delta_{jk}\Phi_{,l}\Phi_{,l}).] \tag{20.29}$$

(*Note:* $t^{jk}_{\text{L-L}}$ as given here is the "stress tensor for a Newtonian gravitational field"; *cf.* exercises 39.5 and 39.6.)

(b) Let the source of the gravitational field be a perfect fluid with

$$T^{\mu\nu} = (\rho + p)u^{\mu}u^{\nu} + pg^{\mu\nu}, \qquad p/\rho \sim v^2 \equiv (dx/dt)^2 \sim |\Phi|.$$

Let the Newtonian potential satisfy the source equation

$$\Phi_{,jj} = 4\pi\rho.$$

Show that the energy of the source is

$$\begin{aligned} P^0 &= \int (T^{00} + t^{00})(-g)\, d^3x \\ &= \int [\rho/\underbrace{(1 - v^2)^{1/2}}_{\text{Lorentz contraction factor}} + \underbrace{\tfrac{1}{2}\rho v^2}_{\text{kinetic energy}} + \underbrace{\tfrac{1}{2}\rho\Phi}_{\text{potential energy}}]\underbrace{(g_{xx}g_{yy}g_{zz})^{1/2}\, dx\, dy\, dz}_{\text{proper volume}} \\ &\quad + \text{higher-order corrections.} \end{aligned} \tag{20.30}$$

(c) Show that the "equations of motion" $T^{\mu\nu}_{\text{L-L eff},\nu} = 0$ reduce to the standard equations (16.3) of Newtonian hydrodynamics.

---

§20.6. EQUATIONS OF MOTION DERIVED FROM THE FIELD EQUATION

Consider the Einstein field equation

$$\boldsymbol{G} = 8\pi \boldsymbol{T} \tag{20.31}$$

under conditions where space is empty of everything except a source-free electromagnetic field:

$$T^{\mu\nu} = \frac{1}{4\pi}\left(F^{\mu\alpha}g_{\alpha\beta}F^{\nu\beta} - \frac{1}{4}g^{\mu\nu}F_{\sigma\tau}F^{\sigma\tau}\right) \tag{20.32}$$

(*cf.* the expression for stress-energy tensor of the electromagnetic field in §5.6). To predict from (20.31) how the geometry changes with time, one has to know how the electromagnetic field changes with time. The field is expressed as the "exterior derivative" of the 4-potential,

$$\boldsymbol{F} = \boldsymbol{dA} \text{ (language of forms)}$$

or

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} \text{ (language of components)}, \tag{20.33}$$

and the time rate of change of the field is governed by the Maxwell equation

$$\boldsymbol{d}{}^*\boldsymbol{F} = 0$$

or

$$F^{\mu\nu}{}_{;\nu} = 0. \tag{20.34}$$

If it seems a fair division of labor for the Maxwell equation to predict the development in time of the Maxwell field and the Einstein equation to do the same for the Einstein field, then it may come as a fresh surprise to discover that the Einstein equation (20.31), plus expression (20.32) for the Maxwell stress-energy, can do both jobs. One does not have to be given the Maxwell "equation of motion" (20.34). One can derive it fresh from (20.31) plus (20.32). The proof proceeds in five steps (see also exercise 3.18 and §5.10). Step one: The Bianchi identity $\boldsymbol{\nabla}\cdot\boldsymbol{G}\equiv 0$ implies conservation of energy-momentum $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$. Step two: Conservation expresses itself in the language of components in the form

Vacuum Maxwell equations derived from Einstein field equation

$$0 = 8\pi T^{\mu\nu}{}_{;\nu} = 2F^{\mu\alpha}{}_{;\nu}g_{\alpha\beta}F^{\nu\beta} + 2F^{\mu\alpha}g_{\alpha\beta}F^{\nu\beta}{}_{;\nu} - g^{\mu\nu}F_{\sigma\tau;\nu}F^{\sigma\tau}. \tag{20.35}$$

Step three: Leaving the middle term unchanged, rearrange the first term so that, like the last term, it carries a factor $F^{\sigma\tau}$. Thus in that first term let the indices $\nu\beta$ of $F^{\nu\beta}$ be replaced in turn by $\sigma\tau$ and by $\tau\sigma$, to subdivide that term into

$$\begin{aligned} &F^{\mu\alpha}{}_{;\sigma}g_{\alpha\tau}F^{\sigma\tau} + F^{\mu\alpha}{}_{;\tau}g_{\alpha\sigma}F^{\tau\sigma} \\ &= (F^{\mu}{}_{\tau;\sigma} - F^{\mu}{}_{\sigma;\tau})F^{\sigma\tau} \\ &= g^{\mu\nu}(F_{\nu\tau;\sigma} + F_{\sigma\nu;\tau})F^{\sigma\tau}. \end{aligned} \tag{20.36}$$

Step four: Combine the first and the last terms in (20.35) to give

$$g^{\mu\nu}(F_{\nu\tau;\sigma} + F_{\sigma\nu;\tau} + F_{\tau\sigma;\nu})F^{\sigma\tau}. \tag{20.37}$$

The indices on the derivatives of the field quantities stand in cyclic order. This circumstance annuls all the terms in the connection coefficients $\Gamma^{\alpha}{}_{\beta\gamma}$ when one writes out the covariant derivatives explicitly. Thus one can replace the covariant derivatives by ordinary derivatives. Moreover, these three derivatives annul one another identically when one substitutes for the fields their expressions (20.33) in terms of the potentials. Consequently, nothing remains in the conservation law (20.35) except the middle term, giving rise to four statements ($\mu = 0, 1, 2, 3$)

$$F^{\mu}{}_{\beta}F^{\beta\nu}{}_{;\nu} = 0 \tag{20.38}$$

about the four quantities ($\beta = 0, 1, 2, 3$)

$$F^{\beta\nu}{}_{;\nu}. \tag{20.39}$$

Step five: The determinant of the coefficients in the four equations (20.38) for the four unknowns (20.39) has the value

$$\begin{vmatrix} F^0{}_0F^0{}_1F^0{}_2F^0{}_3 \\ \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ F^3{}_0F^3{}_1F^3{}_2F^3{}_3 \end{vmatrix} = -(\boldsymbol{E}\cdot\boldsymbol{B})^2 \tag{20.40}$$

(see exercise 20.6, part i). In the generic case, this one function of the four variables (t, x, y, z) vanishes on one or more hypersurfaces; but off any such hypersurface (i.e., at "normal points" in spacetime) it differs from zero. At all normal points, the solution of the four linear equations (20.38) with their nonvanishing determinant gives identically zero for the four unknowns (20.39); that is to say, Maxwell's "equations of motion"

$$F^{\beta\nu}{}_{;\nu} = 0$$

are fulfilled and must be fulfilled as a straight consequence of Einstein's field equation (20.31)—plus expression 20.32 for the stress-energy tensor. Special cases admit counterexamples (see exercise 20.8); but in the generic case one need not invoke Maxwell's equations of motion; one can deduce them from the Einstein field equation.

Lorentz force equation derived from the Einstein field equation

Turn from the dynamics of the Maxwell field itself to the dynamics of a charged particle moving under the influence of the Maxwell field. Make no more appeal to outside providence for the Lorentz equation of motion than for the Maxwell equation of motion. Instead, to generate the Lorentz equation call once more on the Einstein field equation or, more directly, on its consequence, the principle of the local conservation of energy-momentum.

Keep track of the world line of the particle from $t = t$ to $t = t + \Delta t$ (Figure 20.1). Generate a "world tube" around this world line. Thus, at each value of the time coordinate t, take the location of the particle as center; construct a sphere of radius ϵ around this center; and note how the successive spheres sweep out the desired world tube. Construct "caps" on this tube at times t and $t + \Delta t$. The two caps, together with the world tube proper, bound a region of spacetime in which energy and momentum can be neither created nor destroyed ("no creation of moment of rotation," in the language of the Bianchi identities, Chapter 15). Therefore the energy-momentum emerging out of the "top" cap has to equal the energy-momentum entering the "bottom" cap, supplemented by the amount of energy-momentum carried in across the world-tube by the Maxwell field. Out of such an analysis, as performed in flat spacetime, one ends up with the Lorentz equation of motion in its elementary form (see Chapters 3 and 4),

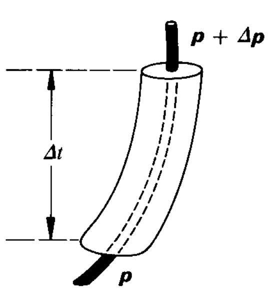


Figure 20.1.
"World tube." The change in the 4-momentum of the particle is governed by the flow of 4-momentum across the boundary of the world tube.

$$d\boldsymbol{p}/d\tau = e\langle \boldsymbol{F}, \boldsymbol{u}\rangle \qquad \text{(language of forms)}$$

or in curved spacetime, the Lorentz equation of motion in covariant form,

$$\nabla_{\boldsymbol{u}}\boldsymbol{p} = m\,\nabla_{\boldsymbol{u}}\boldsymbol{u} = e\langle \boldsymbol{F}, \boldsymbol{u}\rangle \qquad \text{(form language)}$$

or

$$m\left[\frac{d^2x^\alpha}{d\tau^2} + \Gamma^\alpha{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\right] = eF^\alpha{}_\beta\frac{dx^\beta}{d\tau} \quad \text{(component language)}. \tag{20.41}$$

"One ends up with the Lorentz equation of motion"—but only after hurdling problems of principle along the way. One would understand what a particle is if one understood how to do the calculation of balance of energy-momentum with all rigor! Few calculations in all of physics have been done in so many ways by so many leading investigators, from Lorentz and his predecessors to Dirac and Rohrlich [see Teitelboim (1970, 1971) for still further insights]. Among the issues that develop are two that never cease to compel attention. (1) The particle responds according to the Lorentz force law (20.41) to a field. This field is the sum of a contribution from external sources and from the particle itself. How is the field exerted by the particle on itself to be calculated? Insofar as it is not already included in its effects in the "experimental mass" m in (20.41), this force is to be calculated as half the difference between the retarded field and the advanced field caused by that particle (see §36.11 for a more detailed discussion of the corresponding point for an emitter of gravitational radiation). This difference is singularity-free. On the world line, it has the following simple value [valid in general for point particles; valid for finite-sized particles when and only when the particle changes its velocity negligibly compared to the speed of light during the light-travel time across itself—see, e.g., Burke (1970)]

A particle acted on by its own electromagnetic field ("radiation damping")

$$\frac{1}{2}(F_{\text{ret}} - F_{\text{adv}})^{\mu\nu} = \frac{2e}{3}\left(\frac{dx^\mu}{d\tau}\frac{d^3x^\nu}{d\tau^3} - \frac{d^3x^\mu}{d\tau^3}\frac{dx^\nu}{d\tau}\right). \tag{20.42}$$

Every acceptable line of reasoning has always led to expression (20.42). It also represents the field required to reproduce the long-known and thoroughly tested law of radiation damping. (2) "Infinite self-energy." Around a particle at rest, or close to a particle in an arbitrary state of motion, the field is e/r^2 and the field energy is

Infinite self-energy of a point particle

$$(1/8\pi)\int_{r_{\text{min}}}^{\epsilon}(e/r^2)^2 4\pi r^2\,dr = (e^2/2)(r_{\text{min}}^{-1} - \epsilon^{-1}). \tag{20.43}$$

This expression diverges as r_{min} is allowed to go to zero. To hurdle this difficulty, one arranges the calculation of energy balance in such a way that there always appears the sum of this "self-energy" and the "bare mass." The two terms individually are envisaged as "going to infinity" as r_{min} goes to zero; but the sum is identified with the "experimental mass" and is required to remain finite. Of course, no particle is a classical object. A proper calculation of the energy has to be conducted at the quantum level. There it is easier to hide from sight the separate infinities—but they are still present, and promise to remain until the structure of a particle is understood.

Why one is justified to regard equations of motion as consequences of the Einstein field equation

Before one turns from the Maxwell and Lorentz equations of motion to a final example (deriving the geodesic equations of motion for an uncharged particle), is it not time to object to the whole program of "deriving an equation of motion from Einstein's field equation"? First, is it not a pretensious parade of pomposity to say it comes "from Einstein's field equation" (and even more, "from Einstein's field equations") when it really comes from a principle so elementary and long established as the law of conservation of 4-momentum? It cannot be contested that this conservation principle, in historical fact, came before geometrodynamics, just as it came before electrodynamics and before the theories of all other established fields. However, in no theory but Einstein's is this principle incorporated as an identity. Only here does the conservation of energy-momentum appear as a fully automatic consequence of the inner working of the machinery of the world (energy density tied to moment of rotation, and moment of rotation automatically conserved; see Chapter 17). Out of Einstein's theory one can derive the equation of motion of a particle. Out of Maxwell's one cannot. Thus, nothing prevents one from acting on a charge with an "external" force, over and above the Lorentz force, nor from tailoring this force in such a way that the charge follows some prescribed world line ("engine-driven source"). It makes no difficulties whatsoever for Maxwell's equations that one has shifted attention from a world line that follows the Lorentz equation of motion to one that does not. Quite the contrary is true in general relativity. To shift from right world line (geodesic) to wrong world line makes the difference between satisfying Einstein's field equation in the vicinity of that world line and being unable to satisfy Einstein's field equation.

The Maxwell field equations are so constructed that they automatically fulfill and demand the conservation of charge; but not everything has charge. The Einstein field equation is so constructed that it automatically fulfills and demands the conservation of momentum-energy; and everything does have energy. The Maxwell field equations are indifferent to the interposition of an "external" force, because that force in no way threatens the principle of conservation of charge. The Einstein field equation cares about every force, because every force is a medium for the exchange of energy.

Electromagnetism has the motto, "I count all the electric charge that's here." All that bears no charge escapes its gaze.

"I weigh all that's here" is the motto of spacetime curvature. No physical entity escapes this surveillance.

Why, then, is the derivation of the geodesic equation of motion of an object said to be based on "Einstein's geometrodynamic field equation" rather than on "the principle of conservation of 4-momentum"? Because geometry responds by its curvature to mass-energy in every form. Most of all, because geometry outside tells about mass-energy inside, free of all concern about issues of internal structure (violent motions, unknown forces, tortuously curved and even multiply-connected geometry).

If one objection to the plan to derive the equation of motion of a particle "from the field equation" has been disposed of, then the moment has come to deal with the other natural objection: Is there not an inner contradiction in trying to apply to a "particle" (implying idealization to a point) a field equation that deals with the continuum? Answer: There *is* a contradiction in dealing with a point. Therefore do not deal with a point. Do not deal with internal structure at all. Analyze the motion by looking at the geometry outside the object. That geometry provides all the handle one needs to follow the motion.

How one can avoid complexities of particle structure when deriving equations of motion: the "external viewpoint"

Already here one sees the difference from the derivation of the Lorentz equation of motion as sketched out above. There (1) no advantage was taken of geometry outside as indicator of motion inside; (2) a detailed bookkeeping was envisaged of the localization in space of the electromagnetic energy; and (3) this bookkeeping brought up the issue of the internal structure of the particle, which could not be satisfactorily resolved.

Derivation of geodesic motion from Einstein field equation:

Now begin the analysis in the new geometrodynamic spirit. Surrounding "the Schwarzschild zone of influence" of the object, mark out a "buffer zone" (Figure 20.2) that extends out to the region where the "background geometry" begins to depart substantially from flatness. Idealize the geometry in the buffer zone as that of an unchanging source merging asymptotically ("boundary $\mathcal{B}$ of buffer zone") into flat space. It suffices to recall the properties of the spacetime geometry far outside an unchanging (i.e., nonradiating) source (exercise 19.3) to draw the key conclusion: relative to this flat spacetime and regardless of its internal structure, the object remains at rest, or continues to move in a straight line at uniform velocity (conservation of total 4-momentum; §20.5). In other words, it obeys the geodesic equation of motion. If this is the result in a flash, then it is appropriate to go back a step to review it, to find out what it means and what it demands.

(1) derivation in brief

When the object is absent and the background geometry alone has to be considered, then the geodesic is a well-defined mathematical construct. Moreover, Fermi-Walker transport along this geodesic gives a well-defined way to construct a comoving local inertial frame (see §13.6). Relative to this frame, the representative point of the geodesic remains for all time at rest at the origin.

(2) derivation with care

In what way does the presence of the object change this picture? The object possesses an angular momentum, mass quadrupole moments, and higher multipole moments. They interact with the tide-producing accelerations (Riemann curvature) of the background geometry. Depending on the orientation in space of these moments, the interactions drive the object off its geodesic course in one direction or another (see §40.9). These anomalies in the motion go hand in hand with anomalies in the geometry. On and near the ideal mathematical geodesic the metric is Minkowskian. At a point removed from this geodesic by a displacement with Riemann normal coordinates ξ^1, ξ^2, ξ^3 (see §11.6), the metric components differ from their canonical values $(-1, 1, 1, 1)$ by amounts proportional (1) to the squares and products of the ξ^m and (2) to the components of the Riemann curvature tensor (tide-producing acceleration) of the background geometry. These second-order terms produce departures from ideality in the buffer zone, departures that may be described symbolically as of order

Coupling of curvature to particle moments produces deviations from geodesic motion

$$\delta(\text{metric}) \sim r^2 \cdot R \cdot (\text{spherical harmonic of order two}). \tag{20.44}$$

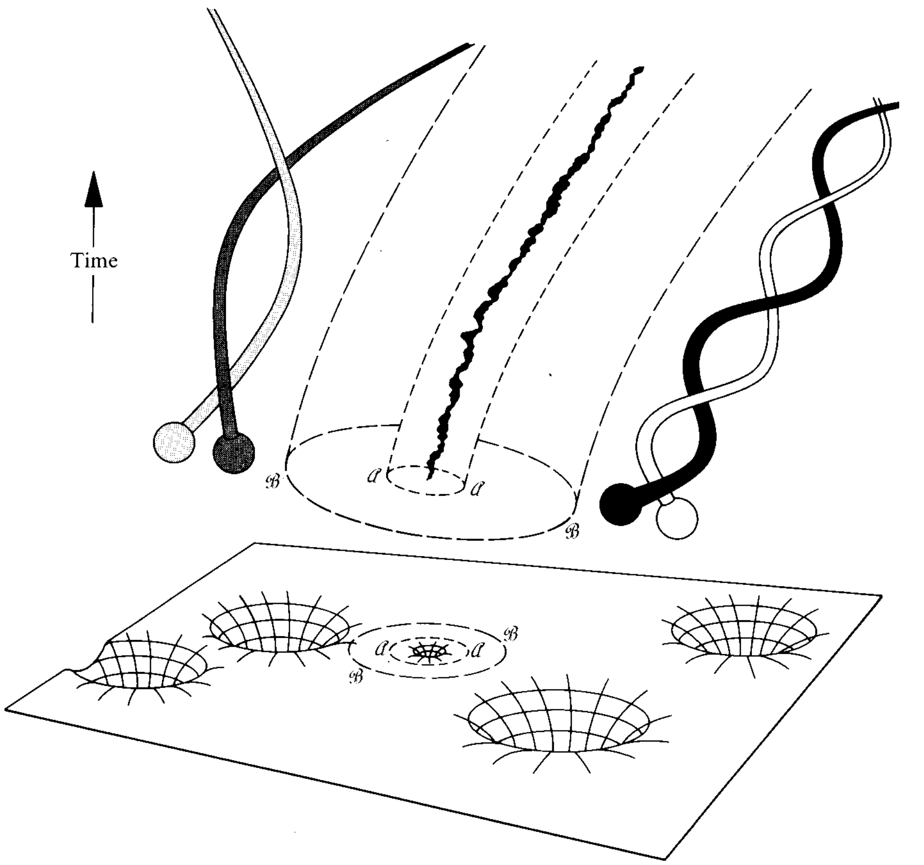


Figure 20.2.
"Buffer zone": the shell of space between 𝒜 and ℬ, where the geometry is appropriately idealized as the spherically symmetric "Schwarzschild geometry" of a localized center of attraction (the object under study) in an asymptotically flat space. Inside 𝒜: the "zone of influence" of the object. In the general case where this object lacks all symmetry, the metric is found to depart more and more from ideal "Schwarzschild character" as the exploration of the geometry is carried inward from 𝒜 (effect of angular momentum of the object on the metric; effect of quadrupole moment; effect of higher moments). Outside ℬ: the "background geometry." As this geometry is explored at greater and greater distances outside ℬ, it is found to depart more and more from flatness (effect of concentrations of mass, gravitational waves, and other geometrodynamics).

Here r is the distance from the geodesic and R is the magnitude of the significant components of the curvature tensor. The object produces not only the standard "Schwarzschild" departure from flatness,

$$\delta(\text{metric}) \sim m/r, \tag{20.45}$$

which by itself (in a flat background) would bring about no departure from geodesic motion, but also correction terms which may be symbolized as

$$\delta(\text{metric}) \sim (S/r^2) \text{ (spherical harmonic of order one)} \tag{20.46}$$

and

$$\delta(\text{metric}) \sim (\mathcal{I}/r^3) \text{ (spherical harmonic of order two)} \tag{20.47}$$

and higher-order terms. Here $S(\text{cm}^2)$ is a typical component of the angular momentum vector or "spin"; $\mathcal{I}(\text{cm}^3)$ is a representative component of the moment of inertia or quadrupole tensor (see Chapter 36 for details), and higher terms have higher-order coefficients.

Coupling of spin to curvature

The tide-producing acceleration generated by the surroundings of the object ("background geometry") acts on the spin of the object with a force of order RS and pulls it away from geodesic motion with an acceleration of the order

$$\text{acceleration } (\text{cm}^{-1}) \sim \frac{R(\text{cm}^{-2})S(\text{cm}^2)}{m(\text{cm})} \tag{20.48}$$

(see exercise 40.8). Otherwise stated, the surrounding and the spin both put warps in the geometry, and these warps conspire together to push the object off track.

The sum of the relevant two perturbations in the metric is qualitatively of the form

$$\delta g \sim r^2 R + S/r^2. \tag{20.49}$$

The sum is least where r has a value of the order

$$r \sim (S/R)^{1/4}, \tag{20.50}$$

and there it has the magnitude

$$\delta g \sim (SR)^{1/2}. \tag{20.51}$$

To "derive the geodesic equation of motion with some preassigned accuracy ϵ" may be defined to mean that the metric in the buffer zone is Minkowskian within the latitude ϵ. In the illustrative example, this means that $(SR)^{1/2}$ is required to be of the order of ϵ or less. Nothing can be done about the value of R because the background curvature R is a feature of the background geometry. One can meet the requirement only by imposing limits on the mass and moments of the object. In the example, where the dominating moment is the angular momentum, one must require that this parameter of the object be less in order of magnitude than the limit

$$S \sim \epsilon^2/R. \tag{20.52}$$

Evidently this and similar conditions on the higher moments are most easily satisfied by demanding that the object have spherical symmetry ($S = 0$, $\mathcal{I} = 0$, higher moments $= 0$). Then the perturbation in the metric, again disregarding angle factors and indices, is qualitatively of the form

$$\delta g \sim r^2 R + m/r, \tag{20.53}$$

and the buffer zone is best designed to bracket the minimizing value of r,

$$r_{\mathscr{A}} \leq [r \sim (m/R)^{1/3}] \leq r_{\mathscr{B}}. \tag{20.54}$$

The departure of the metric from Minkowskian perfection in the buffer zone is of the order

$$\delta g \sim (m^2 R)^{1/3}. \tag{20.55}$$

To achieve any preassigned accuracy ϵ for δg, one must demand that the mass be less than a limit of the order

$$m \sim \epsilon^{3/2}/R^{1/2}. \tag{20.56}$$

The sense in which *no* body can move on a geodesic of spacetime

No object of finite mass moving under the influence of a complex background will admit a buffer zone where the geometry approaches Minkowskian values with arbitrary precision. Therefore it is incorrect to say that such an object follows a geodesic world line. It is meaningless to say that an object of finite rest mass follows a geodesic world line. World line of what? If the object is a black hole, there is no point inside its "horizon" (capture surface; one-way membrane; see Chapters 33 and 34) that is relevant to the physics going on outside. Geodesic world line within what background geometry? It has no sense to speak of a geometry that "lies behind" or is "background to" a black hole.

The sense in which test particles *do* move on geodesics of a background geometry

Turn from one motion of one object in one spacetime to a continuous one-parameter family of spacetimes, with the mass m of the object being the parameter that distinguishes one of these solutions of Einstein's field equation from another. Go to the limit $m = 0$. Then the size of the buffer zone shrinks to zero and the departure of the metric from Minkowskian perfection in the buffer zone also goes to zero. In this limit ("test particle"), it makes sense to say that the object moves in a straight line with uniform velocity in the local inertial frame or, otherwise stated, it pursues a geodesic in the background geometry. Moreover, this background geometry is well-defined: it is the limit of the spacetime geometry as the parameter m goes to zero [see Infeld and Schild (1949)]. In this sense, the geodesic equation of motion follows as an inescapable consequence of Einstein's field equation.

Motion of a charged test particle in curved spacetime

The concept of "background" as limit of a one-parameter family of spacetimes extends itself to the case where the object bears charge as well as mass, and where the surrounding space is endowed with an electromagnetic field. This time the one-parameter family consists of solutions of the combined Einstein-Maxwell equations. The charge-to-mass ratio e/m is fixed. The mass m is again the adjustable parameter. In the limit when m goes to zero, one is left with (1) a background geometry, (2) a background electromagnetic field, and (3) a world line that obeys the general-relativity version of the Lorentz equation of motion in this background as a consequence of the field equations [Chase (1954)]. In contrast, a so-called "unified field theory of gravitation and electromagnetism" that Einstein tentatively put forward at one stage of his thinking, as a conceivable alternative to the combination of his standard 1915 geometrodynamics with Maxwell's standard electrodynamics, has been shown [Callaway (1953)] to lead to the wrong equation of motion for a charged particle. It moves as if uncharged no matter how much charge is piled on its back. If that theory were correct, no cyclotron could operate, no atom could exist, and life itself would be impossible.

References on derivation of equations of motion from Einstein field equation

Thus the ability to yield the correct equation of motion of a particle has today become an added ace in the hand of general relativity. The idea for such a treatment dates back to Einstein and Grommer (1927). Corrections to the geodesic equation of motion arising from interaction between the spin of the object (when it has finite dimensions) and the curvature of the background geometry are treated by Papapetrou (1951) and more completely by Pirani (1956) (see exercise 40.8). A book on the subject exists [Infeld and Plebanski (1960)]. Section 40.9 describes how corrections to geodesic motion affect lunar and planetary orbits. Some of the problems that arise when the object under study fragments or emits a directional stream of radiation, and unresolved issues of principle, are discussed by Wheeler (1961).

When one turns from the limit of infinitesimal mass to an object of finite mass, no simpler situation presents itself then a system of uncharged black holes (Chapter 33). Everything about the motion of these objects follows from an application of the source-free Einstein equation $\boldsymbol{G} = 0$ to the region of spacetime outside the horizons (see Chapter 34) of the several objects. The theory of motion is then geometrodynamics and nothing but geometrodynamics.

Quantum mechanical limitations on the derivation

It has to be emphasized that all the considerations on motion in this section are carried out in the context of classical theory. In the real world of quantum physics, the geometry everywhere experiences unavoidable, natural, zero-point fluctuations (Chapter 43). The calculated local curvatures associated with these fluctuations at the Planck scale of distances [$L = (\hbar G/c^3)^{1/2} = 1.6 \times 10^{-33}$ cm] are enormous [$R \sim 1/L^2 \sim 0.4 \times 10^{66}$ cm^{-2}] compared to the curvature produced on much larger scales by any familiar object (electron or star). No detailed analysis of the interaction of these two curvatures has ever been made. Such an analysis would define a smoothed-out average of the geometry over regions larger than the local quantum fluctuations. With respect to this average geometry, the object will follow geodesic motion: this is the expectation that no one has ever seen any reason to question—but that no one has proved.

EXERCISES

Exercise 20.6. SIMPLE FEATURES OF THE ELECTROMAGNETIC FIELD AND ITS STRESS-ENERGY TENSOR

(a) Show that the "scalar" $-1/2F_{\alpha\beta}F^{\alpha\beta}$ (invariant with respect to coordinate transformations) and the "pseudoscalar" $1/4F_{\alpha\beta}{}^*F^{\alpha\beta}$ (reproduces itself under a coordinate transformation up to a $\pm$ sign, according as the sign of the Jacobian of the transformation is positive

2

or negative) have in any local inertial frame the values $\boldsymbol{E}^2 - \boldsymbol{B}^2$ and $\boldsymbol{E} \cdot \boldsymbol{B}$, respectively ("the two Lorentz invariants" of the electromagnetic field).

(b) Show that the Poynting flux $(\boldsymbol{E} \times \boldsymbol{B})/4\pi$ is less in magnitude than the energy density $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$, save for the exceptional case where both Lorentz invariants of the field vanish (case where the field is locally "null").

(c) A charged pith ball is located a small distance from the North Pole of a bar magnet. Draw the pattern of electric and magnetic lines of force, indicating where the electromagnetic field is "null" in character. Is it legitimate to say that a "null field" is a "radiation field"?

(d) A plane wave is traveling in the z-direction. Show that the corresponding electromagnetic field is everywhere null.

(e) Show that the superposition of two monochromatic plane waves traveling in different directions is null on at most a set of points of measure zero.

(f) In the "generic case" where the field $(\boldsymbol{E}, \boldsymbol{B})$ at the point of interest is not null, show that the Poynting flux is reduced to zero by viewing the field from a local inertial frame that is traveling in the direction of $\boldsymbol{E} \times \boldsymbol{B}$ with a velocity

$$v = \tanh \alpha, \tag{20.57}$$

where the velocity parameter α is given by the formula

$$\tanh 2\alpha = \frac{(\text{Poynting flux})}{(\text{energy density})} = \frac{2|\boldsymbol{E} \times \boldsymbol{B}|}{\boldsymbol{E}^2 + \boldsymbol{B}^2}. \tag{20.58}$$

(g) Show that all components of the electric and magnetic field in this new frame can be taken to be zero except E_x and B_x.

(h) Show that the 4×4 determinant built out of the components of the field in mixed representation, $F_\alpha{}^\beta$, is invariant with respect to general coordinate transformations. (Hint: Use the theorem that the determinant of the product of three matrices is equal to the product of the determinants of those three matrices.)

(i) Show that this determinant has the value $-(\boldsymbol{E} \cdot \boldsymbol{B})^2$ by evaluating it in the special local inertial frame of (f).

(j) Show that in this special frame the Maxwell stress-energy tensor has the form

$$\|T^\mu{}_\nu\| = \frac{E_x{}^2 + B_x{}^2}{8\pi} \begin{Vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{Vmatrix} \tag{20.59}$$

(Faraday tension along the lines of force; Faraday pressure at right angles to the lines of force).

(k) In the other case, where the field is locally null, show that one can always find a local inertial frame in which the field has the form $\boldsymbol{E} = (0, F, 0)$, $\boldsymbol{B} = (0, 0, F)$ and the stress-energy tensor has the value

$$\|T^\mu{}_\nu\| = \frac{F^2}{4\pi} \begin{Vmatrix} -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix} \quad (\mu \text{ for row, } \nu \text{ for column}). \tag{20.60}$$

(l) Regardless of whether the electromagnetic field is or is not null, show that the Maxwell stress-energy tensor has zero trace, $T^\mu{}_\mu = 0$, and that its square is a multiple of the unit tensor,

$$\begin{aligned} T^\mu{}_\alpha T^\alpha{}_\nu &= \frac{\delta^\mu{}_\nu}{(8\pi)^2}[(\boldsymbol{E}^2 - \boldsymbol{B}^2)^2 + (2\boldsymbol{E} \cdot \boldsymbol{B})^2] \\ &= \frac{\delta^\mu{}_\nu}{(8\pi)^2}[(\boldsymbol{E}^2 + \boldsymbol{B}^2)^2 - (2\boldsymbol{E} \times \boldsymbol{B})^2]. \end{aligned} \tag{20.61}$$

Exercise 20.7. **THE STRESS-ENERGY TENSOR DETERMINES THE ELECTROMAGNETIC FIELD EXCEPT FOR ITS COMPLEXION**

(a) Given a non-zero symmetric 4×4 tensor $T^{\mu\nu}$ which has zero trace $T^{\mu}{}_{\mu} = 0$ and whose square is a multiple, $M^4/(8\pi)^2$, of the unit matrix, show that, according as this multiple is zero ("null case") or positive, the tensor can be transformed to the form (20.60) or (20.59) by a suitable rotation in 3-space or by a suitable choice of local inertial frame, respectively.

(b) In the generic (non-null) case in the frame in question, show that $T^{\mu\nu}$ is the Maxwell tensor of the "extremal electromagnetic field" $\xi_{\mu\nu}$ with components

$$\begin{aligned} \boldsymbol{E}^{(\text{extremal})} &= (M, 0, 0), \\ \boldsymbol{B}^{(\text{extremal})} &= (0, 0, 0). \end{aligned} \tag{20.62}$$

Show that it is also the Maxwell tensor of the "dual extremal field" $*\xi_{\mu\nu}$ with components

$$\begin{aligned} *\boldsymbol{E}^{(\text{extremal})} &= (0, 0, 0), \\ *\boldsymbol{B}^{(\text{extremal})} &= (M, 0, 0). \end{aligned} \tag{20.63}$$

(c) Recalling that the duality operation * applied twice to an antisymmetric second-rank tensor (2-form) in four-dimensional space leads back to the negative of that tensor, show that the operator $e^{*\alpha}$ ("duality rotation") has the value

$$e^{*\alpha} = (\cos\alpha) + (\sin\alpha)*. \tag{20.64}$$

(d) Show that the most general electromagnetic field which will reproduce the non-null tensor $T^{\mu\nu}$ in the frame in question, and therefore in any coordinate system, is

$$F_{\mu\nu} = e^{*\alpha}\xi_{\mu\nu}. \tag{20.65}$$

(e) Derive a corresponding result for the null case. [The field $F_{\mu\nu}$ defined in the one frame and therefore in every coordinate system by (d) and (e) is known as the "Maxwell square root" of $T^{\mu\nu}$; $\xi_{\mu\nu}$ is known as the "extremal Maxwell square root" of $T^{\mu\nu}$; and the angle α is called the "complexion of the electromagnetic field." See Misner and Wheeler (1957); see also Boxes 20.1 and 20.2, adapted from that paper.]

Box 20.1 **CONTRAST BETWEEN PROPER LORENTZ TRANSFORMATION AND DUALITY ROTATION**

Quantity	*General proper Lorentz transformation*	*Duality rotation*
Components of the Maxwell stress-energy tensor or the "Maxwell square" of the field $\boldsymbol{F}$	Transformed	Unchanged
The invariants $\boldsymbol{E}^2 - \boldsymbol{B}^2$ and $(\boldsymbol{E}\cdot\boldsymbol{B})^2$	Unchanged	Transformed
The combination $[(\boldsymbol{E}^2 - \boldsymbol{B}^2)^2 + (2\boldsymbol{E}\cdot\boldsymbol{B})^2] = [(\boldsymbol{E}^2 + \boldsymbol{B}^2)^2 - (2\boldsymbol{E}\times\boldsymbol{B})^2]$	Unchanged	Unchanged

2

Box 20.2 TRANSFORMATION OF THE GENERIC (NON-NULL) ELECTROMAGNETIC FIELD TENSOR $\boldsymbol{F} = (\boldsymbol{E}, \boldsymbol{B})$ IN A LOCAL INERTIAL FRAME

Field values	*At start*	*After simplifying duality rotation*
At start	$\boldsymbol{E}, \boldsymbol{B}$	$\boldsymbol{E}$ and $\boldsymbol{B}$ perpendicular, and $\boldsymbol{E}$ greater than $\boldsymbol{B}$
After simplifying Lorentz transformation	$\boldsymbol{E}$ and $\boldsymbol{B}$ parallel to each other and parallel to x-axis	$\boldsymbol{E}$ parallel to x-axis and $\boldsymbol{B} = 0$

Exercise 20.8. THE MAXWELL EQUATIONS CANNOT BE DERIVED FROM THE LAW OF CONSERVATION OF STRESS-ENERGY WHEN $(\boldsymbol{E} \cdot \boldsymbol{B}) = 0$ OVER AN EXTENDED REGION

Supply a counter-example to the idea that the Maxwell equations,

$$F^{\mu\nu}{}_{;\nu} \doteq 0,$$

follow from the Einstein equation; or, more precisely, show that (1) the condition that the Maxwell stress-energy tensor should have a vanishing divergence plus (2) the condition that this Maxwell field is the curl of a 4-potential A_μ can both be satisfied, while yet the stated Maxwell equations are violated. [Hint: It simplifies the analysis without obscuring the main point to consider the problem in the context of flat spacetime. Refer to the paper of Teitelboim (1970) for the decomposition of the retarded field of an arbitrarily accelerated charge into two parts, of which the second, there called $F^{\mu\nu}{}_{II}$, meets the stated requirements, and has everywhere off the worldline $(\boldsymbol{E} \cdot \boldsymbol{B}) = 0$, but does not satisfy the cited Maxwell equations.]

Exercise 20.9. EQUATION OF MOTION OF A SCALAR FIELD AS CONSEQUENCE OF THE EINSTEIN FIELD EQUATION

The stress-energy tensor of a massless scalar field is taken to be

$$T_{\mu\nu} = (1/4\pi)(\phi_{,\mu}\phi_{,\nu} - 1/2 g_{\mu\nu}\phi_{,\alpha}\phi^{,\alpha}). \tag{20.66}$$

Derive the equation of motion of this scalar field from Einstein's field equation.

2

CHAPTER 21

VARIATIONAL PRINCIPLE AND INITIAL-VALUE DATA

Whenever any action occurs in nature, the quantity of action employed by this change is the least possible.

PIERRE MOREAU DE MAUPERTUIS (1746)

In the theory of gravitation, as in all other branches of theoretical physics, a mathematically correct statement of a problem must be determinate to the extent allowed by the nature of the problem; if possible, it must ensure the uniqueness of its solution.

VLADIMIR ALEXANDROVITCH FOCK (1959)

Things are as they are because they were as they were.

THOMAS GOLD (1972)

Calculemus

G. W. LEIBNIZ

§21.1. DYNAMICS REQUIRES INITIAL-VALUE DATA

This chapter is entirely Track 2. No earlier Track-2 material is needed as preparation for it, but Chapters 9–11 and 13–15 will be helpful. It is needed as preparation for Box 30.1 (mixmaster universe) and for Chapters 42 and 43.

No plan for predicting the dynamics of geometry could be at the same time more mistaken and more right than this: "Give the distribution of mass-energy; then solve Einstein's second-order equation,

$$\boldsymbol{G} = 8\pi \boldsymbol{T}, \tag{21.1}$$

for the geometry." Give the distribution of mass-energy in spacetime and solve for the spacetime geometry? No. Give the fields that generate mass-energy, and their

To Karel Kuchař, Claudio Teitelboim, and James York go warm thanks for their collaboration in the preparation of this chapter, and for permission to draw on the lecture notes of K. K. and to quote results of K. K. [especially exercise 21.10] and of J. Y. [especially equations (21.87), (21.88), and (21.152)] prior to publication elsewhere.

time-rates of change, and give 3-geometry of space and its time-rate of change, all at one time, and solve for the 4-geometry of spacetime at that one time? Yes. And only then let one's equations for geometrodynamics and field dynamics go on to predict for all time, in and by themselves, needing no further prescriptions from outside (needing only work!), both the spacetime geometry and the flow of mass-energy throughout this spacetime. This, in brief, is the built-in "plan" of geometrodynamics, the plan spelled out in more detail in this chapter.

Give initial data, predict geometry

Contest the plan. Point out that the art of solving any coupled set of equations lies in separating the unknowns from what is known or to be prescribed. Insist that this separation is already made in (21.1). On the right already stands the source of curvature. On the left already stands the receptacle of curvature in the form of what one wants to know, the metric coefficients, twice differentiated. Claim therefore that one has nothing to do except to go ahead and solve these equations for the metric coefficients. However, in analyzing the structure of the equations to greater depth [see Cartan (1922a) for the rationale of analyzing a coupled set of partial differential equations], one discovers that one can only make the split between "the source and the receptacle" in the right way when one has first recognized the still more important split between "the initial-value data and the future." Thus—to summarize the results before doing the analysis—four of the ten components of Einstein's law connect the curvature of space here and now with the distribution of mass-energy here and now, and the other six equations tell how the geometry as thus determined then proceeds to evolve.

Four of ten components of Einstein equation are conditions on initial-value data

In determining what are appropriate initial-value data to give, one discovers no guide more useful than the Hilbert variational principle,

$$I = \int \mathcal{L}\, d^4x = \int L(-g)^{1/2}\, d^4x \underset{\substack{\uparrow \\ \text{[exercise 8.16]}}}{=} \int L\, d(\text{proper 4-volume}) = \text{extremum} \qquad (21.2)$$

or the Arnowitt-Deser-Misner ("ADM") variant of it (§21.6) and generalizations thereof by Kuchař (§21.9). Out of this principle one can recognize most directly what one must hold fixed at the limits (on an initial spacelike hypersurface and on a final spacelike hypersurface) as one varies the geometry (§21.2) throughout the spacetime "filling of this sandwich," if one is to have a well-defined extremum problem.

The Lagrange function L (scalar function) or the Lagrangian density $\mathcal{L} = (-g)^{1/2}L$ (quantity to be integrated over coordinate volume) is built of geometry alone, when one deals with curved empty space, but normally fields are present as well, and contribute also to the Lagrangian; thus,

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\text{geom}} + \mathcal{L}_{\text{field}} = (-g)^{1/2}L; \\ L &= L_{\text{geom}} + L_{\text{field}}. \end{aligned} \qquad (21.3)$$

The variation of the field Lagrangian with respect to the typical metric coefficient proves to be, of all ways, the one most convenient for generating (that is, for calculating) the corresponding component of the symmetric stress-energy tensor of the field (§21.3).

New view of stress-energy tensor

A computer, allowing for the effect of this field on the geometry and computing ahead from instant to instant the evolution of the metric with time, imposes its own ordering on the events of spacetime. In effect, it slices spacetime into a great number of spacelike slices. It finds it most convenient (§21.4) to do separate bookkeeping on (1) the 3-geometry of the individual slices and (2) the relation between one such slice and the next, as expressed in a "lapse function" N and a 3-vector "shift function" N_i.

The 3-geometry internal to the individual slice or "simultaneity" defines in and by itself the three-dimensional Riemannian curvature intrinsic to this hypersurface; but for a complete account of this hypersurface one must know also the extrinsic curvature (§21.5) telling how this hypersurface is curved with respect to the enveloping four-dimensional spacetime manifold.

In terms of the space-plus-time split of the 4-geometry, the action principle of Hilbert takes a simple and useful form (§21.6).

In the most elementary example of the application of an action principle in mechanics, where one writes

$$I = \int_{x',t'}^{x,t} L(dx/dt, x, t)\, dt \tag{21.4}$$

and extremizes the integral, one already knows that the resultant "dynamic path length" or "dynamic phase" or "action,"

$$S(x, t) = I_{\text{extremum}}, \tag{21.5}$$

is an important quantity, not least because it gives (up to a factor $\hbar$) the phase of the quantum-mechanical wave function. Moreover, the rate of change of this action function with position is what one calls momentum,

$$p = \partial S(x, t)/\partial x; \tag{21.6}$$

and the (negative of the) rate of change with time gives energy (Figure 21.1),

$$E = -\partial S(x, t)/\partial t; \tag{21.7}$$

and the relation between these two features of a system of wave crests,

Hamiltonian as a dispersion relation

$$E = H(p, x), \tag{21.8}$$

call it "dispersion relation" or call it what one will, is the central topic of mechanics.

When dealing with the dynamics of geometry in the Arnowitt-Deser-Misner formulation,* one finds it convenient to think of the specified quantities as being

* *Historical remark.* No one knew until recently what coordinate-free geometric-physical quantity *really* is fixed at limits in the Hilbert-Palatini variational principle. In his pioneering work on the Hamiltonian formulation of general relativity, Dirac paid no particular attention to any variational principle. He had to generalize the Hamiltonian formalism to accommodate it to general relativity, introducing "first- and second-class constraints" and generalizations of the Poisson brackets of classical mechanics. The work of Arnowitt, Deser, and Misner, by contrast, took the variational principle as the foundation for the whole treatment, even though they too did not ask what it is that is fixed at limits in the sense of

Figure 21.1.
Momentum and (the negative of the) energy viewed as rate of change of "dynamic phase" or "action,"

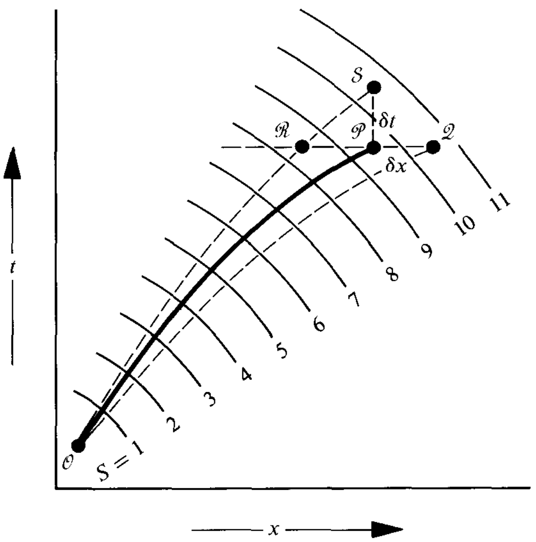
$$S(x, t) = I_{\text{extremum}}(x, t) = \begin{pmatrix}\text{extremum}\\ \text{value of}\end{pmatrix} \int_{x',t'}^{x,t} L(x, \dot{x}, t)\, dt, \qquad (1)$$

with respect to position and time; thus,

$$\delta S = p\, \delta x - E\, \delta t. \qquad (2)$$

The variation of the integral I with respect to changes of the history along the way, $\delta x(t)$, is already zero by reason of the optimization of the history; so the only change that takes place is

$$\begin{aligned}\delta S = \delta I_{\text{extremum}} &= L(x, \dot{x}, t)\, \delta t + \int_{x',t'}^{x+\Delta x,t} \delta L\, dt\\ &= L\, \delta t + \int_{x',t'}^{x+\Delta x,t} \left(\frac{\partial L}{\partial \dot{x}}\delta\dot{x} + \frac{\partial L}{\partial x}\delta x\right) dt\\ &= L\, \delta t + \frac{\partial L}{\partial \dot{x}}\Delta x + \int_{x',t'}^{x+\Delta x,t} \underbrace{\left(\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}}\right)}_{\text{zero by reason of extremization}} \delta x\, dt. \qquad (3)\end{aligned}$$



When one contemplates only a change δx in the coordinates (x, t) of the end point (change of history from $\mathcal{OP}$ to $\mathcal{OQ}$), one has $\Delta x = \delta x$. When one makes only a change δt in the end point (change of history from $\mathcal{OP}$ to $\mathcal{OS}$), one has $\Delta x =$ (indicator of change from $\mathcal{P}$ to $\mathcal{R}$) $= -\dot{x}\, \delta t$. For the general variation of the final point, one thus has $\Delta x = \delta x - \dot{x}\, \delta t$ and

$$\delta S = \frac{\partial L}{\partial \dot{x}}\delta x - \left(\dot{x}\frac{\partial L}{\partial \dot{x}} - L\right)\delta t. \qquad (4)$$

One concludes that the "dispersion relation" is obtained by taking the relations [compare (2) and (4)]

$$\begin{pmatrix}\text{rate of change of}\\ \text{dynamic phase}\\ \text{with position}\end{pmatrix} = (\text{momentum}) = p = \frac{\partial L(x, \dot{x}, t)}{\partial \dot{x}} \qquad (5)$$

and

$$-\begin{pmatrix}\text{rate of change of}\\ \text{dynamic phase}\\ \text{with time}\end{pmatrix} = (\text{energy}) = E = \dot{x}\frac{\partial L}{\partial \dot{x}} - L, \qquad (6)$$

and eliminating $\dot{x}$ from them [solve (5) for $\dot{x}$ and substitute that value of $\dot{x}$ into (6)]; thus

$$E = H(p, x, t) \qquad (7)$$

or

$$-\frac{\partial S}{\partial t} = H\left(\frac{\partial S}{\partial x}, x, t\right). \qquad (8)$$

Every feature of this elementary analysis has its analog in geometrodynamics.

---

a coordinate-free geometric-physical quantity. The great payoff of this work was recognition of the lapse and shift functions of equation (21.40) as Lagrange multipliers, the coefficients of which gave directly and simply Dirac's constraints. They did not succeed in arriving at a natural and simple time-coordinate, but that goal has in the meantime been achieved in the "extrinsic time" of Kuchař and York (§21.11). However, the Arnowitt-Deser Misner approach opened the door to the "intrinsic time" of Sharp, Baierlein, and Wheeler, where 3-geometry is fixed at limits, and 3-geometry is the carrier of information about time; and this led directly to Wheeler's "superspace version" of the treatment of Arnowitt, Deser, and Misner.

the 3-geometry ${}^{(3)}\mathfrak{G}$ of the initial spacelike hypersurface and the 3-geometry ${}^{(3)}\mathfrak{G}$ of the final spacelike hypersurface. One envisages the action integral as extremized with respect to the choice of the spacetime that fills the "sandwich" between these two faces. If one has thus determined the spacetime, one has automatically by that very act determined the separation in proper time of the two hypersurfaces. There is no additional time-variable to be brought in or considered. The one concept ${}^{(3)}\mathfrak{G}$ thus takes the place in geometrodynamics of the two quantities x, t of particle dynamics. The action S that there depended on x and t here depends on the 3-geometry of the face of the sandwich; thus,

Action viewed as dependent on 3-geometry

$$S = S({}^{(3)}\mathfrak{G}). \tag{21.9}$$

A change in the 3-geometry changes the action. The amount of the change in action per elementary change in 3-geometry defines the "field momentum" π^{ij}_{true} conjugate to the geometrodynamic field coordinate g_{ij}, according to the formula

$$\delta S = \int \pi^{ij}_{\text{true}}\, \delta g_{ij}\, d^3x. \tag{21.10}$$

Comparing this equation out of the Arnowitt, Deser, and Misner (ADM) canonical formulation of geometrodynamics (§21.7) with the expression for change of action with change of endpoint in elementary mechanics,

$$\delta S = p\, \delta x - E\, \delta t, \tag{21.11}$$

one might at first think that something is awry, there being no obvious reference to time in (21.10). However, the 3-geometry is itself automatically the carrier of information about time; and (21.10) is complete. Moreover, with no "time" variable other than the information that ${}^{(3)}\mathfrak{G}$ itself already carries about time, there is also no "energy." Thus the "dispersion relation" that connects the rates of change of action with respect to the several changes that one can make in the "field coordinates" or 3-geometry takes the form

$$\mathcal{H}(\pi^{ij}, g_{mn}) = 0, \tag{21.12}$$

with the E-term of (21.8) equal to zero (details in §21.7). All the content of Einstein's general relativity can be extracted from this one Hamiltonian, or "super-Hamiltonian," to give it a more appropriate name [see DeWitt (1967a), pp. 1113–1118, for an account of the contributions of Dirac, of Arnowitt, Deser, and Misner, and of others to the Hamiltonian formulation of geometrodynamics; and see §21.7 and subsequent sections of this chapter for the meaning and payoffs of this formulation].

Hamiltonian versus super-Hamiltonian

The difference between a Hamiltonian and a super-Hamiltonian [see, for example, Kramers (1957)] shows nowhere more clearly than in the problem of a charged particle moving in flat space under the influence of the field derived from the electromagnetic 4-potential, $A_\mu(x^\alpha)$. The Hamiltonian treatment derives the equation of motion from the action principle,

$$0 = \delta I = \delta \int \left[p_i \frac{dx^i}{dt} - H(p_j, x^k, t) \right] dt$$

with

$$H = -\frac{e}{c}\phi + \left[m^2 + \eta^{ij}\left(p_i + \frac{e}{c}A_i\right)\left(p_j + \frac{e}{c}A_j\right)\right]^{1/2}.$$

The super-Hamiltonian analysis gets the equations of motion from the action principle

$$0 = \delta I' = \delta \int \left[p_\mu \frac{dx^\mu}{d\lambda} - \mathcal{H}(p_\alpha, x^\beta)\right] d\lambda.$$

Here the super-Hamiltonian is given by the expression

$$\mathcal{H}(p_\alpha, x^\beta) = \frac{1}{2}\left[m^2 + \eta^{\mu\nu}\left(p_\mu + \frac{e}{c}A_\mu\right)\left(p_\nu + \frac{e}{c}A_\nu\right)\right].$$

The variational principle gives Hamilton's equations for the rates of change

$$dx^\alpha/d\lambda = \partial\mathcal{H}/\partial p_\alpha$$

and

$$dp_\beta/d\lambda = -\partial\mathcal{H}/\partial x^\beta.$$

From these equations, one discovers that $\mathcal{H}$ itself must be a constant, independent of the time-like parameter λ. The value of this constant has to be imposed as an initial condition, $\mathcal{H} = 0$ ("specification of particle mass"), thereafter maintained by the Hamiltonian equations themselves. This vanishing of $\mathcal{H}$ in no way kills the partial derivatives,

$$\partial\mathcal{H}/\partial p_\alpha \qquad \text{and} \qquad -\partial\mathcal{H}/\partial x^\beta,$$

that enter Hamilton's equations for the rates of change,

$$dx^\alpha/d\lambda \qquad \text{and} \qquad dp_\beta/d\lambda.$$

Whether derived in the one formalism or the other, the equations of motion are equivalent, but the covariance shows more clearly in the formalism of the super-Hamiltonian, and similarly in general relativity.

Granted values of the "field coordinates" $g_{ij}(x, y, z)$ (${}^{(3)}\mathcal{G}$) and field momenta $\pi^{ij}_{\text{true}}(x, y, z) = \delta S/\delta g_{ij}$ compatible with (21.12), one has what are called "compatible initial-value data on an initial spacelike hypersurface." One can proceed as described in §21.8 to integrate ahead in time step by step from one spacelike hypersurface to another and another, and construct the whole 4-geometry. Here one is dealing with what in mathematical terminology are hyperbolic differential equations that have the character of a wave equation.

Dynamic evolution of geometry

In contrast, one deals with elliptic differential equations that have the character of a Poisson potential equation when one undertakes in the first place to construct the needed initial-value data (§21.9). In the analysis of these elliptic equations, it proves helpful to distinguish in the 3-geometry between (1) the part of the metric that determines relative lengths at a point, which is to say angles ("the conformal part of the metric") and (2) the common multiplicative factor that enters all the components of the g_{ij} at a point to determine the absolute scale of lengths at that point. This breakdown of the 3-geometry into two parts provides a particularly simple way to deal with two special initial-value problems known as the time-symmetric and time-antisymmetric initial-value problems (§21.10).

Another choice of what to fix at boundary hypersurface: conformal part of 3-geometry plus extrinsic time

The ADM formalism is today in course of development as summarized in §21.11. In Wheeler's (1968a) "superspace" form, the ADM treatment takes the 3-geometry to be fixed on each of the bounding spacelike hypersurfaces. In contrast, York (§21.11) goes back to the original Hilbert action principle, and discovers what *it* takes to be fixed on each of the bounding spacelike hypersurfaces. The appropriate data turn out to be the "conformal part of the 3-geometry" plus something closely related to what Kuchař (1971a and 1972) calls the "extrinsic time." The contrast between Wheeler's approach and the Kuchař-York approach shows particularly clearly when one (1) deals with a flat spacetime manifold, (2) takes a flat spacelike section through this spacetime, and then (3) introduces a slight bump on this slice, of height ϵ. The 3-geometry intrinsic to this deformed slice differs from Euclidean geometry only to the second order in ϵ. Therefore to read back from the full 3-geometry to the time ("the forward advance of the bump") requires in this case an operation something like extracting a square root. In contrast, the Kuchař-York treatment deals with the "extrinsic curvature" of the slice, something proportional to the first power of ϵ, and therefore provides what is in some ways a more convenient measure of time [see especially Kuchař (1971) for the construction of "extrinsic time" for arbitrarily strong cylindrical gravitational waves; see also Box 30.1 on "time" as variously defined in "mixmaster cosmology"]. York shows that the time-variable is most conveniently identified with the variable "dynamically conjugate to the conformal factor in the 3-geometry."

Mach updated: mass-energy there governs inertia here

The initial-value problem of geometrodynamics can be formulated either in the language of Wheeler or in the language of Kuchař and York. In either formulation (§21.9 or §21.11) it throws light on what one ought properly today to understand by Mach's principle (§21.12). That principle meant to Mach that the "acceleration" dealt with in Newtonian mechanics could have a meaning only if it was acceleration with respect to the fixed stars or to something equally well-defined. It guided Einstein to general relativity. Today it is summarized in the principle that "mass-energy there governs inertia here," and is given mathematical expression in the initial-value equations.

The analysis of the initial-value problem connected past and future across a spacelike hypersurface. In contrast, one encounters a hypersurface that accommodates a timelike vector when one deals (§21.13) with the junction conditions between one solution of Einstein's field equation (say, the Friedmann geometry interior to a spherical cloud of dust of uniform density) and another (say, the Schwarzschild geometry exterior to this cloud of dust). Section 21.13, and the chapter, terminate with notes on gravitational shock waves and the characteristic initial-value problem (the statement of initial-value data on a light cone, for example).

§21.2. THE HILBERT ACTION PRINCIPLE AND THE PALATINI METHOD OF VARIATION

Variational principle the simplest route to Einstein's equation

Five days before Einstein presented his geometrodynamic law in its final and now standard form, Hilbert, animated by Einstein's earlier work, independently discovered (1915a) how to formulate this law as the consequence of the simplest action principle of the form (21.2–21.3) that one can imagine:

$$L_{\text{geom}} = (1/16\pi)^{(4)}R. \tag{21.13}$$

(Replace $1/16\pi$ by $c^3/16\pi G$ when going from the present geometric units to conventional units; or divide by $\hbar \sim L^{*2}$ to convert from dynamic phase, with the units of action, to actual phase of a wave function, with the units of radians). Here ${}^{(4)}R$ is the four-dimensional scalar curvature invariant, as spelled out in Box 8.4.

This action principle contains second derivatives of the metric coefficients. In contrast, the action principle for mechanics contains only first derivatives of the dynamic variables; and similarly only derivatives of the type $\partial A_\alpha/\partial x^\beta$ appear in the action principle for electrodynamics. Therefore one might also have expected only first derivatives, of the form $\partial g_{\mu\nu}/\partial x^\gamma$, in the action principle here. However, no scalar invariant lets itself be constructed out of these first derivatives. Thus, to be an invariant, L_{geom} has to have a value independent of the choice of coordinate system. But in the neighborhood of a point, one can always so choose a coordinate system that all first derivatives of the $g_{\mu\nu}$ vanish. Apart from a constant, there is no scalar invariant that can be built homogeneously out of the metric coefficients and their first derivatives.

When one turns from first derivatives to second derivatives, one has all twenty distinct components of the curvature tensor to work with. Expressed in a local inertial frame, these twenty components are arbitrary to the extent of the six parameters of a local Lorentz transformation. There are thus $20 - 6 = 14$ independent local features of the curvature ("curvature invariants") that are coordinate-independent, any one of which one could imagine employing in the action principle. However, ${}^{(4)}R$ is the only one of these 14 quantities that is linear in the second derivatives of the metric coefficients. Any choice of invariant other than Hilbert's complicates the geometrodynamic law, and destroys the simple correspondence with the Newtonian theory of gravity (Chapter 17).

Scalar curvature invariant the only natural choice

Hilbert originally conceived of the independently adjustable functions of x, y, z, t in the variational principle as being the ten distinct components of the metric tensor in contravariant representation, $g^{\mu\nu}$. Later Palatini (1919) discovered a simpler and more instructive listing of the independently adjustable functions: not the ten $g^{\mu\nu}$ alone, but the ten $g^{\mu\nu}$ plus the forty $\Gamma^\alpha_{\mu\nu}$ of the affine connection.

To give up the standard formula for the connection Γ in terms of the metric g and let Γ "flap in the breeze" is not a new kind of enterprise in mathematical physics. Even in the simplest problem of mechanics, one can give up the standard formula for the momentum p in terms of a time-derivative of the coordinate x and also let

Idea of varying coordinate and momentum independently

p "flap in the breeze." Then $x(t)$ and $p(t)$ become two independently adjustable functions in a new variational principle,

$$I = \int_{x',t'}^{x,t} \left[p(t)\frac{dx(t)}{dt} - H(p(t), x(t), t) \right] dt = \text{extremum.} \tag{21.14}$$

Happily, out of the extremization with respect to choice of the function $p(t)$, one recovers the standard formula for the momentum in terms of the velocity. The extremization with respect to choice of the other function, $x(t)$, gives the equation of motion just as does the more elementary variational analysis of Euler and Lagrange, where $x(t)$ is the sole adjustable function. A further analysis of this equivalence between the two kinds of variational principles in particle mechanics appears in Box 21.1. In that box, one also sees the two kinds of variational principle as applied to electrodynamics.

To express the Hilbert variational principle in terms of the $\Gamma^\lambda_{\mu\nu}$ and $g^{\alpha\beta}$ regarded as the primordial functions of t, x, y, z, note that the Lagrangian density is

$$L_{\text{geom}}(-g)^{1/2} = (1/16\pi)^{(4)}R(-g)^{1/2} = (1/16\pi)g^{\alpha\beta}R_{\alpha\beta}(-g)^{1/2}. \tag{21.15}$$

Here, as in any spacetime manifold with an affine connection, one has (Chapter 14)

$$R_{\alpha\beta} = R^\lambda{}_{\alpha\lambda\beta}, \tag{21.16}$$

where

$$R^\lambda{}_{\alpha\mu\beta} = \partial\Gamma^\lambda{}_{\alpha\beta}/\partial x^\mu - \partial\Gamma^\lambda{}_{\alpha\mu}/\partial x^\beta + \Gamma^\lambda_{\sigma\mu}\Gamma^\sigma_{\alpha\beta} - \Gamma^\lambda_{\sigma\beta}\Gamma^\sigma_{\alpha\mu}, \tag{21.17}$$

and every Γ is given in advance (in a coordinate frame) as symmetric in its two lower indices. In order that the integral I of (21.2–21.3) should be an extremum, one requires that the variation in I caused by changes both in the $g^{\mu\nu}$ and in the Γ's should vanish; thus,

$$0 = \delta I = (1/16\pi)\int \delta[g^{\alpha\beta}R_{\alpha\beta}(-g)^{1/2}]\, d^4x + \int \delta[L_{\text{field}}(-g)^{1/2}]\, d^4x. \tag{21.18}$$

Consider now the variations of the individual factors in the first and second integrals in (21.18). The variation of the first factor is trivial, $\delta g^{\alpha\beta}$. In the variation of the second factor, $R_{\alpha\beta}$, changes in the $g^{\alpha\beta}$ play no part; only changes in the Γ's appear. Moreover, the variation $\delta\Gamma^\lambda_{\alpha\beta}$ is a tensor even though $\Gamma^\lambda_{\alpha\beta}$ itself is not. Thus in the transformation formula

Variation of connection is a tensor

$$\Gamma^{\bar\gamma}{}_{\bar\alpha\bar\beta} = \left[\Gamma^\lambda_{\sigma\tau}\frac{\partial x^\sigma}{\partial x^{\bar\alpha}}\frac{\partial x^\tau}{\partial x^{\bar\beta}} + \frac{\partial^2 x^\lambda}{\partial x^{\bar\alpha}\partial x^{\bar\beta}} \right] \frac{\partial x^{\bar\gamma}}{\partial x^\lambda}, \tag{21.19}$$

the last term destroys the tensor character of any set of $\Gamma^\lambda_{\sigma\tau}$ individually, but subtracts out in the difference $\delta\Gamma^\lambda_{\sigma\tau}$ between two alternative sets of Γ's. Note that the variation $\delta R^\lambda{}_{\alpha\mu\beta}$ of the typical component of the curvature tensor consists of two terms of

(*continued on page 500*)

Box 21.1 RATE OF CHANGE OF ACTION WITH DYNAMIC COORDINATE (= "MOMENTUM") AND WITH TIME, AND THE DISPERSION RELATION (= "HAMILTONIAN") THAT CONNECTS THEM IN PARTICLE MECHANICS AND IN ELECTRODYNAMICS

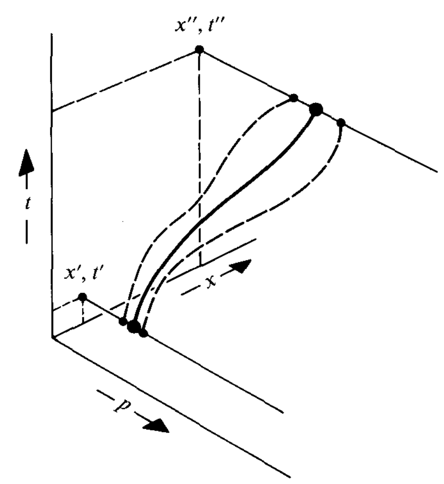
A. PROLOG ON THE PARTICLE-MECHANICS ANALOG OF THE PALATINI METHOD

In particle mechanics, one considers the history $x = x(t)$ to be adjustable between the end points (x', t') and (x, t) and varies it to extremize the integral $I = \int L(x, \dot{x}, t)\, dt$ taken between these two limits.

Expressed in terms of coordinates and momenta (see Figure 21.1), the integral has the form

$$I = \int [p\dot{x} - H(p, x, t)]\, dt, \tag{1}$$

where $x(t)$ is again the function to be varied and p is only an abbreviation for a certain function of x and $\dot{x}$; thus, $p = \partial L(x, \dot{x}, t)/\partial \dot{x}$. Viewed in this way, the variation, $\delta p(t)$, of the momentum is governed by, and is only a reflection of, the variation $\delta x(t)$.



1. Momentum Treated as Independently Variable

There miraculously exists, however, quite another way to view the problem (see inset). One can regard $x(t)$ and $p(t)$ as two quite uncorrelated and independently adjustable functions. One abandons the formula $p = \partial L(x, \dot{x}, t)/\partial \dot{x}$, only to recover it, or the equivalent of it, from the new "independent-coordinate-and-momentum version" of the variation principle.

The variation of (1), as defined and calculated in this new way, becomes

$$\delta I = p\,\delta x \Big|_{x',t'}^{x'',t''} + \int_{x',t'}^{x'',t''} \left[\left(\dot{x} - \frac{\partial H}{\partial p}\right)\delta p + \left(-\dot{p} - \frac{\partial H}{\partial x}\right)\delta x\right] dt. \tag{2}$$

Demand that the coefficient of δp vanish and have the sought-for new version,

$$\dot{x} = \frac{\partial H(p, x, t)}{\partial p}$$

of the old relation, $p = \partial L(x, \dot{x}, t)/\partial \dot{x}$, between momentum and velocity. The vanishing of the coefficient of δx gives the other Hamilton equation,

$$\dot{p} = -\frac{\partial H(p, x, t)}{\partial x}, \tag{3}$$

equivalent in content to the original Lagrange equation of motion,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0. \tag{4}$$

That $p(t)$ in this double variable conception is—before the extremization!—a function of time quite separate from and independent of the function $x(t)$ shows nowhere more clearly than in the circumstance that $p(t)$ has no end point conditions imposed on it, whereas x' and x'' are specified. Thus not only is the shape of the history subject to adjustment in x, p, t space in the course of achieving the extremum, but even the end points are subject to being slid along the two indicated lines in the inset, like beads on a wire.

2. Action as Tool for Finding Dispersion Relation

Denote by $S(x, t)$ the "action," or extremal value of I, for the classical history that starts with (x', t') and ends at (x, t) ($= \hbar$ times phase of de Broglie wave). To change the end points to $(x + \delta x, t)$ makes the change in action

$$\delta S = p\,\delta x. \tag{5}$$

Thus momentum is "rate of change of action with dynamic coordinate."

To change the end point to

$$(x + \delta x,\ t + \delta t) = ([x + \dot{x}\,\delta t] + [\delta x - \dot{x}\,\delta t],\ t + \delta t) \tag{6}$$

makes the change in action

$$\delta S = p[\delta x - \dot{x}\,\delta t] + L\,\delta t = p\,\delta x - H\,\delta t. \tag{7}$$

Thus the Hamiltonian is the negative of "the rate of change of action with time."

In terms of the Hamiltonian $H = H(p, x)$, the "dispersion relation" for de Broglie waves becomes

$$-\frac{\partial S}{\partial t} = H\left(\frac{\partial S}{\partial x}, x\right). \tag{8}$$

In the derivation of this dispersion relation, one can profitably short-cut all talk of $p(t)$ and $x(t)$ as independently variable quantities, and derive the result in hardly more than one step from the definition $I = \int L(x, \dot{x}, t)\, dt$. Similarly in electrodynamics.

The remainder of this box best follows a first perusal of Chapter 21.

B. ANALOG OF THE PALATINI METHOD IN ELECTRODYNAMICS

In source-free electrodynamics, one considers as given two spacelike hypersurfaces S' and S'', and the magnetic fields-as-a-function-of-position in each, B' and B'' (this second field will later be written without the $''$ superscript to simplify the notation). To be varied is an integral extended over the region of spacetime between the two hypersurfaces,

$$I_{\text{Maxwell}} \equiv \int \mathcal{L}_{\text{Maxwell}}\, d^4x = -\frac{1}{16\pi} \int F^{\mu\nu} F_{\mu\nu} (-g)^{1/2}\, d^4x. \qquad (9)$$

1. Variation of Field on Hypersurface and Variation of Location of Hypersurface are Cleanly Separated Concepts in Electromagnetism

The electromagnetic field $\boldsymbol{F}$ is the physically relevant quantity in electromagnetism (compare the 3-geometry in geometrodynamics). By contrast, the 4-potential $\boldsymbol{A}$ has no direct physical significance. A change of gauge in the potentials,

$$A_\mu = A_{\mu_{\text{new}}} + \partial\lambda/\partial x^\mu$$

leaves unchanged the field components

$$F_{\mu\nu} = \partial A_\nu/\partial x^\mu - \partial A_\mu/\partial x^\nu$$

(compare the coordinate transformation that changes the $g_{\mu\nu}$ while leaving unchanged the ${}^{(3)}\mathcal{G}$). The variation of the fields within the body of the sandwich is nevertheless expressed most conveniently in terms of the effect of changes δA_μ in the potentials.

One also wants to see how the action integral is influenced by changes in the location of the upper spacelike hypersurface ("many-fingered time"). Think of the point of the hypersurface that is presently endowed with coordinates $x, y, z, t(x, y, z)$ as being displaced to $x, y, z, t + \delta t(x, y, z)$. Now renounce this use of a privileged coordinate system. Describe the displacement of the simultaneity in terms of a 4-vector $\delta\boldsymbol{n}$ (not a unit 4-vector) normal to the hypersurface Σ. The element of 4-volume $\delta\Omega$ included between the original upper face of the sandwich and the new upper face, that had in the privileged coordinate system the form $(-g)^{1/2}\, \delta t(x, y, z)\, d^3x$, in the notation of Chapter 20 becomes

$$\delta\Omega = \delta n^\mu\, d^3\Sigma_\mu = (\delta\boldsymbol{n} \cdot d^3\boldsymbol{\Sigma}), \qquad (10)$$

where the element of surface $d^3\Sigma_\mu$ already includes the previously listed factor $(-g)^{1/2}$.

Box 21.1 ***(continued)***

Counting together the influence of changes in the field values on the upper hypersurface and changes in the location of that hypersurface, one has

$$\delta S = \delta I_{\text{extremal}} = -(1/16\pi)\int_{\text{upper }\Sigma} F^{\mu\nu}F_{\mu\nu}(\delta \boldsymbol{n}\cdot d^3\boldsymbol{\Sigma})$$

$$+ (1/4\pi)\int_{\text{upper }\Sigma} F^{\mu\nu}\underbrace{\Delta A_\mu}_{\substack{\text{replace by}\\ \text{its equivalent}\\ (\delta A_\mu - \delta n^\alpha A_{\mu;\alpha})}} d^3\Sigma_\nu \tag{11}$$

$$+ (1/4\pi)\int_{\text{4-volume}} \underbrace{F^{\mu\nu}{}_{;\nu}}_{\substack{\text{has to vanish}\\ \text{because integral has}\\ \text{been extremized}}} \delta A_\mu (-g)^{1/2}\, d^4x.$$

Simplify this expression by arranging the coordinates so that the hypersurface shall be a hypersurface of constant t, and so that lines of constant x, y, z shall be normal to this hypersurface. Then it follows that the element of volume on that hypersurface contains a single nonvanishing component, $d^3\Sigma_0 = (-g)^{1/2}\, d^3x$. The antisymmetry of the field quantity $F^{0\nu}$ in its two indices requires that ν be a spacelike label, $i = 1, 2, 3$. The variation of the action becomes

$$\delta S = \int\left[\frac{(-g)^{1/2}F^{i0}}{4\pi}\delta A_i - \underbrace{\left\{\underbrace{\frac{(-g)^{1/2}F^{i0}}{4\pi}}A_{i;0} - \mathcal{L}_{\text{Maxwell}}\right\}\delta t}_{\text{add and subtract }\left\{\frac{(-g)^{1/2}F^{i0}}{4\pi}A_0\,\delta t\right\}_{,i}}\right] d^3x. \tag{12}$$

2. Meaning of Field "Momentum" in Electrodynamics

Identify this expression with the quantity

$$\delta S = \int \pi^i_{EM}\,\delta A_i\, d^3x - \int \mathcal{H}\,\delta\Omega, \tag{13}$$

where

$$\pi^i_{EM} = \frac{\delta S}{\delta A_i} = \left(\begin{array}{l}\text{“density of electromagnetic}\\ \text{momentum dynamically canon-}\\ \text{ically conjugate to } A_i\text{”}\end{array}\right) = \frac{(-g)^{1/2}F^{i0}}{4\pi} = -\frac{\mathcal{E}^i}{4\pi} \tag{14}$$

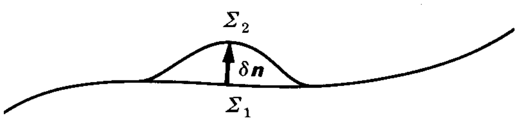
is a simple multiple of the electric field and where

$$\mathscr{H} = -\frac{\delta S}{\delta \Omega} = \begin{pmatrix}\text{“density of}\\ \text{electromagnetic}\\ \text{Hamiltonian”}\end{pmatrix} = (1/16\pi)[F^{\mu\nu}F_{\mu\nu} + 4F^{i0}(A_{i;0} - A_{0;i})] \qquad (15)$$
$$= (1/8\pi)(\boldsymbol{E}^2 + \boldsymbol{B}^2).$$

The concept of dynamic Hamiltonian density agrees with the usual concept of density of electromagnetic energy, despite the very different context in which the two quantities are derived and used. However, the canonical momentum π^i_{EM} has nothing directly whatsoever to do with the density of electromagnetic momentum as defined, for example, by the Poynting vector, despite the confusing similarity in the standard names for the two quantities. Note that there is no term δA_0 in (13); that is, $\pi^0_{EM} \equiv 0$.

3. Bubble Differentiation

The “bubble differentiation” with respect to “many-fingered time” that appears in (15) was first introduced by Tomonaga (1946). One thinks of a spacelike hypersurface Σ_1, a magnetic field $\boldsymbol{B}$ defined as a function of position on this hypersurface (by an observer on a world line normal to this hypersurface), and a prescription S that carries one from this information to a single number, the action. (Divided by $\hbar$, this action gives the phase of the “wave function” or “probability amplitude” for the occurrence of this particular distribution of field values over this particular hypersurface.) One goes to a second hypersurface Σ_2 (see inset), which is identical with Σ_1, except in the immediate vicinity of a given point. Take a distribution of field values over Σ_2 that is identical with the original distribution over Σ_1, “identity of location” being defined by means of the normal. Evaluate the difference, δS, in the value of the dynamic phase or action in the two cases. Divide this difference by the amount of proper 4-volume $\delta\Omega = \int(\delta\boldsymbol{n} \cdot d^3\boldsymbol{\Sigma})$ contained in the “bubble” between the two hypersurfaces. Take the quotient, evaluate it in the limit in which the size of the bubble goes to zero, and in this way get the “bubble-time derivative,” $\delta S/\delta\Omega$, of the action.



Box 21.1 (*continued*)

What does it mean to say that the action, S, besides depending on the hypersurface, Σ, depends also on the distribution of the magnetic field, B, over that hypersurface? The action depends on the physical quantity, $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$, not on the prephysical quantity, $\boldsymbol{A}$. Thus a change in gauge $\delta A_i = \partial\lambda/\partial x^i$, cannot make any change in S. On the other hand, the calculated value of the change in S for this alteration in $\boldsymbol{A}$ is

$$\begin{aligned} \delta(\text{action}) = \delta S &= \int \frac{\delta S}{\delta A_i}\,\delta A_i\,d^3x \\ &= \int \frac{\delta S}{\delta A_i}\frac{\partial\lambda}{\partial x^i}\,d^3x = -\int \left(\frac{\delta S}{\delta A_i}\right)_{,i} \lambda(x, y, z)\,d^3x. \end{aligned} \tag{16}$$

In order that there shall be no dependence of action on gauge, it follows that this expression must vanish for arbitrary $\lambda(x, y, z)$, a result only possible if $S(\Sigma, \boldsymbol{B}) = S$(hypersurface, field on hypersurface) satisfies the identity

$$\left(\frac{\delta S}{\delta A_i}\right)_{,i} = \pi^{i}_{EM,i} = -(1/4\pi)\mathscr{E}^{i}{}_{,i} = 0. \tag{17}$$

4. Hamilton-Jacobi "Propagation Law" for Electrodynamics

The "dispersion relation" or "Hamilton-Jacobi equation" for electromagnetism relates (1) the changes of the "dynamic phase" or "action" brought about by alterations in the dynamic variables A_i (the generalization of the x of particle dynamics) with (2) the changes brought about by alterations in many-fingered time (the generalization of the single time t of particle dynamics); thus (15) translates into

$$-\frac{\delta S}{\delta \Omega} = \frac{(4\pi)^2}{8\pi}\left(\frac{\delta S}{\delta A}\right)^2 + \frac{1}{(8\pi)}(\boldsymbol{\nabla} \times \boldsymbol{A})^2 \tag{18}$$

C. DISPERSION RELATIONS FOR GEOMETRODYNAMICS AND ELECTRODYNAMICS COMPARED AND CONTRASTED

Geometrodynamics possesses a direct analog of equation (17) ("action depends on no information carried by the vector potential A except the magnetic field $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$"), in an equation that says the action depends on no information carried by the metric g_{ij} on the "upper face of the sandwich" except the 3-geometry there, ${}^{(3)}\mathscr{G}$. It also possesses a direct analog of equation (18) ("dynamic equation for the propagation of the action") with this one difference: in electrodynamics the field variable $\boldsymbol{B}$ and the many-fingered time are distinct in character, whereas in geometrodynamics the "field" and the "many-fingered time" can be regarded as two aspects of one and the same ${}^{(3)}\mathscr{G}$:

D. ACTION PRINCIPLE AND DISPERSION RELATION ARE ROOTED IN THE QUANTUM PRINCIPLE; FEYNMAN'S PRINCIPLE OF THE DEMOCRATIC EQUALITY OF ALL HISTORIES

For more on action principles in physics, see for example Mercier (1953), Lanczos (1970), and Yourgrau and Mandelstam (1968).

Newton (1687) in the first page of the preface to the first edition of his *Principia* notes that "The description of right lines . . . , upon which geometry is founded, belongs to mechanics. Geometry does not teach us to draw these lines, but requires them to be drawn."

Newton's remark is also a question. Mechanics moves a particle along a straight line, but what is the machinery by which mechanics accomplishes this miracle? The quantum principle gives the answer. The particle moves along the straight line only by not moving along the straight line. In effect it "feels out" every conceivable world line that leads from the start, (x', t'), to the point of detection, (x'', t''), "compares" one with another, and takes the extremal world line. How does it accomplish this miracle?

The particle is governed by a "probability amplitude to transit from (x', t') to (x'', t'')." This amplitude or "propagator," $\langle x'', t''|x', t'\rangle$, is the democratic sum with equal weight of contributions from every world line that leads from start to finish; thus,

$$\langle x'', t''|x', t'\rangle = N\int e^{iI_H/\hbar}\, \mathcal{D}x. \tag{15}$$

Here N is a normalization factor, the same for all histories.

$\mathcal{D}x$ is the "volume element" for the sum over histories. For a "skeleton history" defined by giving x_n at $t_n = t_0 + n\,\Delta t$, one has $\mathcal{D}x$ equal, up to a multiplicative constant, to $dx_1\, dx_2 \ldots dx_N$. When the history is defined by the Fourier coefficients in such an expression as

$$x(t) = \frac{x'(t'' - t) + x''(t - t')}{(t'' - t')} + \sum_n a_n \sin n\pi \frac{(t - t')}{(t'' - t')}, \tag{16}$$

the volume element, again up to a multiplicative factor, is $da_1\, da_2 \ldots$.

Destructive interference in effect wipes out the contribution to the transition probability from histories that differ significantly from the "extremal history" or "classical history." Histories that are near that extremal history, on the other hand, contribute constructively, and for a simple reason: a small departure of the first order from the classical history brings about a change in phase which is only of the second order in the departure.

In this elementary example, one sees illustrated why it is that extremal principles play such a large part in classical dynamics. They remind one that all classical physics rests on a foundation of quantum physics. The central ideas are (1) the principle

Box 21.1 (*continued*)

of superposition of probability amplitudes, (2) constructive and destructive interference, (3) the "democracy of all histories," and (4) the probability amplitude associated with a history H is $e^{iI_H/\hbar}$, apart from a normalizing factor that is a multiplicative constant, ∞ .

For more on the democracy of histories and the sum over histories see Feynman (1942, 1948, 1949, 1951, and 1955), and the book of Feynman and Hibbs (1965); also Hibbs (1951), Morette (1951), Choquard (1955), Polkinghorne (1955), Fujiwara (1962), and the survey and literature references in Kursunoglu (1962); also reports of Dempster (1963) and Symanzik (1963). This outlook has been applied by many workers to discuss the quantum formulation of geometrodynamics, the first being Misner (1957) and one of the latest being Faddeev (1971).

the form $\delta\Gamma^{\lambda}{}_{\alpha\beta,\mu}$ and four terms of the form $\Gamma\,\delta\Gamma$ (indices being dropped for simplicity). One coordinate system is as good as another in dealing with a tensor. Therefore pick a coordinate system in which all the Γ's vanish at the point under study. The terms $\Gamma\,\delta\Gamma$ drop out. In this coordinate system, the variation of the curvature is expressed in terms of first derivatives of quantities like $\delta\Gamma^{\lambda}_{\alpha\beta}$. One then need only replace the ordinary derivatives by covariant derivatives to obtain a formula correct in any coordinate system,

$$\delta R^{\lambda}_{\alpha\mu\beta} = \delta\Gamma^{\lambda}_{\alpha\beta;\mu} - \delta\Gamma^{\lambda}_{\alpha\mu;\beta}, \tag{21.20}$$

along with its contraction,

$$\delta R_{\alpha\beta} = \delta\Gamma^{\lambda}_{\alpha\beta;\lambda} - \delta\Gamma^{\lambda}_{\alpha\lambda;\beta}. \tag{21.21}$$

The third factor that appears in the variation principle is $(-g)^{1/2}$. Its variation (exercise 21.1) is

$$\delta(-g)^{1/2} = -\frac{1}{2}(-g)^{1/2}g_{\mu\nu}\,\delta g^{\mu\nu}. \tag{21.22}$$

The other integrand, the Lagrange density L_{field}, will depend on the fields present and their derivatives, but will be assumed to contain the metric only as $g^{\mu\nu}$ itself, never in the form of any derivatives of $g^{\mu\nu}$.

In order for an extremum to exist, the following expression has to vanish:

$$\begin{aligned}(1/16\pi)\int\Big[\Big(R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R\Big)\delta g^{\alpha\beta} + g^{\alpha\beta}(\delta\Gamma^{\lambda}_{\alpha\beta;\lambda} - \delta\Gamma^{\lambda}_{\alpha\lambda;\beta})\Big](-g)^{1/2}\,d^4x \\ + \int\Big(\frac{\delta L_{\text{field}}}{\delta g^{\alpha\beta}} - \frac{1}{2}g_{\alpha\beta}L_{\text{field}}\Big)\delta g^{\alpha\beta}(-g)^{1/2}\,d^4x = 0\end{aligned} \tag{21.23}$$

Focus attention on the term in (21.23) that contains the variations of Γ,

$$(1/16\pi)\int g^{\alpha\beta}(\delta\Gamma^{\lambda}_{\alpha\beta;\lambda} - \delta\Gamma^{\lambda}_{\alpha\lambda;\beta})(-g)^{1/2}\,d^4x,$$

and integrate by parts to eliminate the derivatives of the $\delta\Gamma$. To prepare the way for this integration, introduce the concept of *tensor density*, a notational device widely applied in general relativity. The concept of tensor density aims at economy. Without this concept, one will treat the tensor

Concept of tensor density

$$\epsilon_{\mu\alpha\beta\gamma} = (-g)^{1/2}[\mu\alpha\beta\gamma]$$

(see exercise 3.13) as having $4^4 = 256$ components, and its covariant derivative as having $4^5 = 1{,}024$ components, of which one is

$$\begin{aligned}\epsilon_{0123;\rho} &= \partial(-g)^{1/2}/\partial x^{\rho}\epsilon_{[0123]} - \Gamma^{\sigma}_{0\rho}\epsilon_{\sigma123} - \Gamma^{\sigma}_{1\rho}\epsilon_{0\sigma23}\\ &\quad - \Gamma^{\sigma}_{2\rho}\epsilon_{01\sigma3} - \Gamma^{\sigma}_{3\rho}\epsilon_{012\sigma}\\ &= [(-g)^{1/2}{}_{,\rho} - \Gamma^{\sigma}_{\sigma\rho}(-g)^{1/2}][0123].\end{aligned}$$

The symbol $[\alpha\beta\gamma\delta]$, with values $(0, -1, +1)$, introduces what is largely excess baggage, doing mere bookkeeping on alternating indices. Drop this unhandiness. Introduce instead the non-tensor $(-g)^{1/2}$ and *define* for it the law of covariant differentiation,

$$(-g)^{1/2}{}_{;\rho} = (-g)^{1/2}{}_{,\rho} - \Gamma^{\sigma}_{\sigma\rho}(-g)^{1/2}. \tag{21.24}$$

These four components take the place of the 1,024 components and communicate all the important information that was in them.

Associated with the vector j_μ is the vector density

$$\mathfrak{j}_\mu = (-g)^{1/2}j_\mu;$$

with the tensor $T_{\mu\nu}$, the tensor density

$$\mathfrak{T}_{\mu\nu} = (-g)^{1/2}T_{\mu\nu};$$

and so on; the German gothic letter is a standard indicator for the presence of the factor $(-g)^{1/2}$. On some occasions (see, for example, §21.11) it is convenient to multiply the components of a tensor with a power of $(-g)^{1/2}$ other than 1. According to the value of the exponent, the resulting assemblage of components is then called a tensor density of this or that *weight*.

The law of differentiation of an ordinary or standard tensor density formed from a tensor of arbitrary order,

$$\mathfrak{A}^{\cdot\cdot\cdot}_{\cdot\cdot} = (-g)^{1/2}A^{\cdot\cdot\cdot}_{\cdot\cdot},$$

is

$$\begin{aligned}(\mathfrak{A}^{\cdot\cdot\cdot}_{\cdot\cdot})_{;\rho} = (\mathfrak{A}^{\cdot\cdot\cdot}_{\cdot\cdot})_{,\rho} &+ (\text{standard } \Gamma^{\cdot}_{\cdot\cdot} \text{ terms of a standard covariant}\\ &\text{derivative multiplied into } \mathfrak{A}^{\cdot\cdot\cdot}_{\cdot\cdot}) - (\mathfrak{A}^{\cdot\cdot\cdot}_{\cdot\cdot})\Gamma^{\sigma}_{\sigma\rho}.\end{aligned}$$

The covariant derivative of a product is the sum of two terms: the covariant derivative of the first, times the second, plus the first times the covariant derivative of the second.

Now return to the integral to be evaluated. Combine the factors $g^{\alpha\beta}$ and $(-g)^{1/2}$ into the tensor density $\mathfrak{g}^{\alpha\beta}$. Integrate covariantly by parts, as justified by the rule for the covariant derivative of a product. Get a "term at limits," plus the integral

$$-(1/16\pi)\int (\mathfrak{g}^{\alpha\beta}{}_{;\lambda} - \delta^\beta_\lambda \mathfrak{g}^{\alpha\gamma}{}_{;\gamma})\, \delta\Gamma^\lambda_{\alpha\beta}\, d^4x.$$

This integral is the only term in the action integral that contains the variations of the Γ's at the "interior points" of interest here. For the integral to be an extremum, the symmetrized coefficient of $\delta\Gamma^\lambda_{\alpha\beta}$ must vanish,

$$\mathfrak{g}^{\alpha\beta}{}_{;\lambda} - \frac{1}{2}\delta^\alpha_\lambda \mathfrak{g}^{\beta\gamma}{}_{;\gamma} - \frac{1}{2}\delta^\beta_\lambda \mathfrak{g}^{\alpha\gamma}{}_{;\gamma} = 0.$$

This set of forty equations for the forty covariant derivative $\mathfrak{g}^{\alpha\beta}{}_{;\lambda}$ has only the zero solution,

$$\mathfrak{g}^{\alpha\beta}{}_{;\lambda} = 0. \tag{21.25}$$

Thus the "density formed from the reciprocal metric tensor" is covariantly constant.

This simple result (1) brings many simple results in its train: the covariant constancy of (2) $(-g)^{1/2}$, (3) $g^{\alpha\beta}$, (4) $g_{\alpha\beta}$, and (5) $\mathfrak{g}_{\alpha\beta}$. Of these, (4) is of special interest here, and (2) is needed in proving it, as follows. Take definition (21.24) for the covariant derivative of $(-g)^{1/2}$, and calculate the ordinary derivative that appears in the first term from exercise 21.1. One encounters in this calculation terms of the form $\partial\mathfrak{g}^{\alpha\beta}/\partial x^\lambda$. Use (21.25) to evaluate them, and end up with the result

$$(-g)^{1/2}{}_{;\lambda} = 0.$$

From this result it follows that the covariant derivative of the $\binom{1}{1}$-tensor density $(-g)^{1/2}\,\delta^\alpha_\gamma$ is also zero. But this tensor density is the product of the tensor density $\mathfrak{g}^{\alpha\beta}$ by the ordinary metric tensor $g_{\beta\gamma}$. In the covariant derivative of this product by x^λ, one already knows that the derivative of the first factor is zero. Therefore the first factor times the derivative of the second must be zero,

$$\mathfrak{g}^{\alpha\beta} g_{\beta\gamma;\lambda} = 0,$$

and from this it follows that

$$g_{\beta\gamma;\lambda} = 0, \tag{21.26}$$

as was to be proven; or, explicitly,

$$\frac{\partial g_{\beta\gamma}}{\partial x^\lambda} - g_{\gamma\sigma}\Gamma^\sigma_{\beta\lambda} - g_{\beta\sigma}\Gamma^\sigma_{\gamma\lambda} = 0.$$

Solve these equations for the Γ's, which up to now have been independent of the $g_{\beta\gamma}$, and end up with the standard equation for the connection coefficients,

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\sigma}(g_{\mu\sigma,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma}), \tag{21.27}$$

as required for Riemannian geometry.

Similarly, equate to zero the coefficient of $\delta g^{\alpha\beta}$ in the variation (21.23), and find all ten components of Einstein's field equation, in the form

$$G_{\alpha\beta} = 8\pi \underbrace{\left(g_{\alpha\beta} L_{\text{field}} - 2\frac{\delta L_{\text{field}}}{\delta g^{\alpha\beta}}\right)}_{\left[\begin{array}{l}\text{identified in §21.3 with}\\ \text{the stress-energy tensor } T_{\alpha\beta}\end{array}\right]}. \tag{21.28}$$

Among variations of the metric, one of the simplest is the change

$$g_{\text{new}\,\mu\nu} = g_{\mu\nu} + \delta g_{\mu\nu} = g_{\mu\nu} + \xi_{\mu;\nu} + \xi_{\nu;\mu} \tag{21.29}$$

brought about by the infinitesimal coordinate transformation

$$x^{\mu}_{\text{new}} = x^{\mu} - \xi^{\mu}. \tag{21.30}$$

Although the metric changes, the 3-geometry does not. It does not matter whether the spacetime geometry that one is dealing with extremizes the action principle or not, whether it is a solution of Einstein's equations or not; the action integral I is a scalar invariant, a number, the value of which depends on the physics but not at all on the system of coordinates in which that physics is expressed. This invariance even obtains for both parts of the action principle individually (I_{geom} and I_{fields}). Therefore neither part will be affected in value by the variation (21.29). In other words, the quantity

Action unaffected by mere change in coordinatization

$$\begin{aligned}\delta I_{\text{geom}} &= (1/16\pi) \int G_{\alpha\beta}(\xi^{\alpha;\beta} + \xi^{\beta;\alpha})(-g)^{1/2}\, d^4x \\ &\underset{\text{[“covariant integration by parts”]}}{=} -(1/8\pi) \int G_{\alpha\beta}{}^{;\beta}\xi^{\alpha}(-g)^{1/2}\, d^4x\end{aligned} \tag{21.31}$$

must vanish whatever the 4-geometry and whatever the change ξ^{α}. In this way, one sees from a new angle the contracted Bianchi identities of Chapter 15,

$$G_{\alpha\beta}{}^{;\beta} = 0. \tag{21.32}$$

The "neutrality" of the action principle with respect to a mere coordinate transformation such as (21.29) shows once again that the variational principle—and with it Einstein's equation—cannot determine the coordinates or the metric, but only the 4-geometry itself.

EXERCISE

Exercise 21.1. VARIATION OF THE DETERMINANT OF THE METRIC TENSOR

Recalling that the change in the value of any determinant is given by multiplying the change in each element of that determinant by its cofactor and adding the resulting products (exercise 5.5) prove that

$$\delta(-g)^{1/2} = \frac{1}{2}(-g)^{1/2} g^{\mu\nu}\,\delta g_{\mu\nu} \qquad \text{and} \qquad \delta(-g)^{1/2} = -\frac{1}{2}(-g)^{1/2} g_{\mu\nu}\,\delta g^{\mu\nu}.$$

Also show that

$$g = \det \|\mathfrak{g}^{\mu\nu}\| \qquad \text{and} \qquad \delta(-g)^{1/2} = +\frac{1}{2} g_{\mu\nu}\,\delta \mathfrak{g}^{\mu\nu}.$$

---

§21.3. MATTER LAGRANGIAN AND STRESS-ENERGY TENSOR

The derivation of Einstein's geometrodynamic law from Hilbert's action principle puts on the righthand side a source term that is derived from the field Lagrangian. In contrast, the derivation of Chapter 17 identified the source term with the stress-energy tensor of the field. For the two derivations to be compatible, the stress-energy tensor must be given by the expression

Lagrangian generates stress-energy tensor

$$T_{\alpha\beta} = -2\frac{\delta L_{\text{field}}}{\delta g^{\alpha\beta}} + g_{\alpha\beta} L_{\text{field}}, \tag{21.33a}$$

or

$$(-g)^{1/2} T^{\alpha\beta} \equiv \mathfrak{T}^{\alpha\beta} = 2\frac{\delta \mathcal{L}_{\text{field}}}{\delta g_{\alpha\beta}}. \tag{21.33b}$$

What are the consequences of this identification?

By the term "Lagrange function of the field" as employed here, one means the Lagrange function of the classical theory as formulated in flat spacetime, with the flat-spacetime metric replaced wherever it appears by the actual metric, and with the "comma-goes-to-semicolon rule" of Chapter 16 applied to all derivatives.

Were one dealing with a general tensorial field, the comma-goes-to-semicolon rule would introduce, in addition to the derivative of the tensorial field with all its indices, a number of Γ's equal to the number of indices. The presence of these Γ's in the field Lagrangian would have unhappy consequences for the Palatini variational procedure described in §21.2. No longer would the Γ's end up given in terms of the metric coefficients by the standard formula (21.27). No longer would the geometry, as derived from the Hilbert-Palatini variation principle, be Riemannian. Then what?

These troublesome issues do not arise in two well-known simple cases, a scalar field and an electromagnetic field. In the one case, the field Lagrangian becomes

$$L_{\text{field}} = (1/8\pi)[-g^{\alpha\beta}(\partial\phi/\partial x^\alpha)(\partial\phi/\partial x^\beta) - m^2\phi^2]. \tag{21.34}$$

Electromagnetism as an example

No connection coefficient comes in; the quantity being differentiated is a scalar. In the other case, the field Lagrangian is built on first derivatives of the 4-potential A_μ. Therefore Γ's should appear, according to the standard rules for covariant differentiation (Box 8.4). However, the derivatives of the A's appear, never alone, but always in an antisymmetric combination where the Γ's cancel, making covariant derivatives equivalent to ordinary derivatives:

$$F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu} = A_{\nu,\mu} - A_{\mu,\nu}. \tag{21.35}$$

In both cases, the differentiations of (21.33) to generate the stress-energy tensor are easily carried out (exercises 21.2 and 21.3) and give the standard expressions already seen [(5.22) and (5.23)] for $T_{\mu\nu}$ in one of these two cases in an earlier chapter.

Contrast to stress-energy tensor of "canonical field theory"

Field theory provides a quite other method to generate a so-called canonical expression for the stress-energy tensor of a field [see, for example, Wentzel (1949)].

By the very manner of construction, such an expression is guaranteed also to satisfy the law of conservation of momentum and energy, and by this circumstance it too becomes useful in certain contexts. However, the canonical tensor is often not symmetric in its two indices, and in such cases violates the law of conservation of angular momentum (see discussion in §5.7). Even when symmetric, it may give a quite different localization of stress and energy than that given by (21.33). Field theory in and by itself is unable to decide between these different pictures of where the field energy is localized. However, direct measurements of the pull of gravitation provide in principle [see, for example, Feynman (1964)] a means to distinguish between alternative prescriptions for the localization of stress-energy, because gravitation responds directly to density of mass-energy and momentum. It is therefore a happy circumstance that the theory of gravity in the variational formulation gives a unique prescription for fixing the stress-energy tensor, a prescription that, besides being symmetric, also automatically satisfies the laws of conservation of momentum and energy (exercises 21.2 and 21.3). [For an early discussion of the symmetrization of the stress-energy tensor, see Rosenfeld (1940) and Belinfante (1940). A more extensive discussion is given by Corson (1953) and Davis (1970), along with extensive references to the literature.]

When one deals with a spinor field, one finds it convenient to take as the quantities to be varied, not the metric coefficients themselves, but the components of a tetrad of orthonormal vectors defined as a tetrad field over all space [see Davis (1970) for discussion and references].

EXERCISES

Exercise 21.2. STRESS-ENERGY TENSOR FOR A SCALAR FIELD

Given the Lagrange function (21.34) of a scalar field, derive the stress-energy tensor for this field. Also write down the field equation for the scalar field that one derives from this Lagrange function (in the general case where the field executes its dynamics within the arena of a curved spacetime). Show that as a consequence of this field equation, the stress-energy tensor satisfies the conservation law, $T_{\alpha\beta}{}^{;\beta} = 0$.

Exercise 21.3. FARADAY-MAXWELL STRESS-ENERGY TENSOR

Given the Lagrangian density $-F_{\mu\nu}F^{\mu\nu}/16\pi$, reexpress it in terms of the variables A_μ and $g^{\mu\nu}$, and by use of (21.33) derive the stress-energy tensor as discussed in §5.6. Also derive from the Lagrange variation principle the field equation $F_{\alpha\beta}{}^{;\beta} = 0$ (curved spacetime, but—for simplicity—a charge-free region of space). As a consequence of this field equation, show that the Faraday-Maxwell stress-energy tensor satisfies the conservation law, $T_{\alpha\beta}{}^{;\beta} = 0$. For a more ambitious project, show that any stress-energy tensor derived from a field Lagrangian by the prescription of equation (21.33) will automatically satisfy the conservation law $T_{\alpha\beta}{}^{;\beta} = 0$.

---

§21.4. SPLITTING SPACETIME INTO SPACE AND TIME

There are many ways to "push forward" many-fingered time and explore spacetime faster here and slower there, or faster there and slower here. However, a computer is most efficiently programmed only when it follows one definite prescription. The

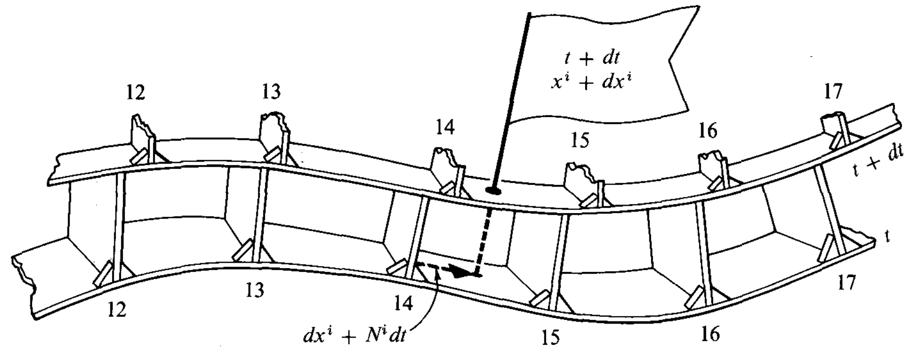


Figure 21.2.
Building two 3-geometries into a thin sandwich 4-geometry, by interposing perpendicular connectors between the two, with preassigned lengths and shifts. What would otherwise be flexible thereupon becomes rigid. The flagged point illustrates equation (21.40).

Slice spacetime to compute spacetime

successive hypersurfaces on which it gives the geometry are most conveniently described by successive values of a time-parameter t. One treats on a different footing the 3-geometries of these hypersurfaces and the 4-geometry that fills in between these laminations.

The slicing of spacetime into a one-parameter family of spacelike hypersurfaces is called for, not only by the analysis of the dynamics along the way, but also by the boundary conditions as they pose themselves in any action principle of the form, "Give the 3-geometries on the two faces of a sandwich of spacetime, and adjust the 4-geometry in between to extremize the action."

Thin sandwich 4-geometry

There is no simpler sandwich to consider than one of infinitesimal thickness (Figure 21.2). Choosing coordinates adapted to the (3 + 1)-space-time split, designate the "lower" (earlier) hypersurface in the diagram as t = constant and the "upper" (later) one as $t + dt$ = constant (names, only names; no direct measure whatsoever of proper time). Compare the two hypersurfaces with two ribbons of steel out of which one wants to construct a rigid structure. To give the geometry on the two ribbons by no means fixes this structure; for that purpose, one needs cross-connectors between the one ribbon and the other. It is not even enough (1) to specify that these connectors are to be welded on perpendicular to the lower ribbon; (2) to specify where each is to be welded; and (3) to give its length. One must in addition tell where each connector joins the upper surface. If the proper distances between tops of the connectors are everywhere shorter than the distances between the bases of the connectors, the double ribbon will have the curve of the cable of a suspension bridge; if everywhere longer, the curve of the arch of a masonry bridge. The data necessary for the construction of the sandwich are thus (1) the metric of the 3-geometry of the lower hypersurface,

$$g_{ij}(t, x, y, z)\, dx^i\, dx^j, \tag{21.36}$$

telling the (distance)^2 between one point in that hypersurface and another; (2) the metric on the upper hypersurface,

$$g_{ij}(t + dt, x, y, z)\, dx^i\, dx^j; \tag{21.37}$$

(3) a formula for the proper length,

$$\begin{pmatrix}\text{lapse of}\\ \text{proper time}\\ \text{between lower}\\ \text{and upper}\\ \text{hypersurface}\end{pmatrix} = \begin{pmatrix}\text{“lapse}\\ \text{function”}\end{pmatrix} dt = N(t, x, y, z)\, dt, \tag{21.38}$$

of the connector that is based on the point (x, y, z) of the lower hypersurface; and (4) a formula for the place on the upper hypersurface,

$$x^i_{\text{upper}}(x^m) = x^i - N^i(t, x, y, z)\, dt, \tag{21.39}$$

where this connector is to be welded. Omit part of this information, and find the structure deprived of rigidity.

Metric of 4-geometry depends on lapse and shift of connectors of the two 3-geometries

The rigidity of the structure of the thin sandwich is most immediately revealed in the definiteness of the 4-geometry of the spacetime filling of the sandwich. Ask for the proper interval ds or $d\tau$ between $x^\alpha = (t, x^i)$ and $x^\alpha + dx^\alpha = (t + dt,\, x^i + dx^i)$. The Pythagorean theorem in its 4-dimensional form

$$ds^2 = \begin{pmatrix}\text{proper distance}\\ \text{in base 3-geometry}\end{pmatrix}^2 - \begin{pmatrix}\text{proper time from}\\ \text{lower to upper 3-geometry}\end{pmatrix}^2$$

yields the result (see Figure 21.2).

$$ds^2 = g_{ij}(dx^i + N^i\, dt)(dx^j + N^j\, dt) - (N\, dt)^2 \tag{21.40}$$

Here as in (21.36) the g_{ij} are the metric coefficients of the 3-geometry, distinguished by their Latin labels from the Greek-indexed components of the 4-metric,

$$ds^2 = {}^{(4)}g_{\alpha\beta}\, dx^\alpha\, dx^\beta, \tag{21.41}$$

Details of the 4-geometry

labeled here with a suffix $^{(4)}$ to reduce the possibility of confusion. Comparing (21.41) and (21.40), one arrives at the following construction of the 4-metric out of the 3-metric and the lapse and shift functions [Arnowitt, Deser, and Misner (1962)]:

$$\begin{Vmatrix} {}^{(4)}g_{00} & {}^{(4)}g_{0k} \\ {}^{(4)}g_{i0} & {}^{(4)}g_{ik} \end{Vmatrix} = \begin{Vmatrix} (N_s N^s - N^2) & N_k \\ N_i & g_{ik} \end{Vmatrix}. \tag{21.42}$$

The welded connectors do the job!

In (21.42), the quantities N^m are the components of the shift in its original primordial contravariant form, whereas the $N_i = g_{im} N^m$ are the covariant components, as calculated within the 3-geometry with the 3-metric. To invert this relation,

$$N^m = g^{ms} N_s \tag{21.43}$$

is to deal with the reciprocal 3-metric, a quantity that has to be distinguished sharply from the reciprocal 4-metric. Thus, the reciprocal 4-metric is

$$\left\| \begin{matrix} {}^{(4)}g^{00} & {}^{(4)}g^{0m} \\ {}^{(4)}g^{k0} & {}^{(4)}g^{km} \end{matrix} \right\| = \left\| \begin{matrix} -(1/N^2) & (N^m/N^2) \\ (N^k/N^2) & (g^{km} - N^k N^m/N^2) \end{matrix} \right\|, \tag{21.44}$$

a result that one checks by calculating out the product

$${}^{(4)}g_{\alpha\beta}\,{}^{(4)}g^{\beta\gamma} = {}^{(4)}\delta_\alpha{}^\gamma$$

according to the standard rules for matrix multiplication.

The volume element has the form

$$(-{}^{(4)}g)^{1/2}\, dx^0\, dx^1\, dx^2\, dx^3 = N g^{1/2}\, dt\, dx^1\, dx^2\, dx^3. \tag{21.45}$$

Welding the connectors to the two steel ribbons, or adding the lapse and shift functions to the 3-metric, by rigidifying the 4-metric, also automatically determines the components of the unit timelike normal vector $\boldsymbol{n}$. The condition of normalization on this 4-vector is most easily formulated by saying that there exists a 1-form, also called $\boldsymbol{n}$ for the sake of convenience, dual to $\boldsymbol{n}$, and such that the product of this vector by this 1-form has the value

$$\langle \boldsymbol{n}, \boldsymbol{n} \rangle = -1. \tag{21.46}$$

This 1-form has the value

$$\boldsymbol{n} = n_\beta\, \boldsymbol{d}x^\beta = -N\, \boldsymbol{d}t + 0 + 0 + 0. \tag{21.47}$$

Only so can this 1-form, this structure of layered surfaces, automatically yield a value of unity, one bong of the bell, when pierced as in Figure 2.4 by a vector that represents an advance of one unit in proper time, regardless of what x, y, and z displacements it also has. Thus the unit timelike normal vector in covariant 1-form representation necessarily has the components

The components of the unit normal

$$n_\beta = (-N, 0, 0, 0) \tag{21.48}$$

Raise the indices via (21.44) to obtain the contravariant components of the same normal, represented as a tangent vector; thus,

$$n^\alpha = [(1/N), -(N^m/N)]. \tag{21.49}$$

This result receives a simple interpretation on inspection of Figure 21.2. Thus the typical "perpendicular connector" in the diagram can be said to have the components

$$(dt, -N^m\, dt)$$

and to have the proper length $d\tau = N\,dt$; so, ratioed down to a vector $\boldsymbol{n}$ of unit proper length, the components are precisely those given by (21.49).

§21.5. INTRINSIC AND EXTRINSIC CURVATURE

The central concept in Einstein's account of gravity is curvature, so it is appropriate to analyze curvature in the language of the (3 + 1)-space-time split. The curvature intrinsic to the 3-geometry of a spacelike hypersurface may be defined and calculated by the same methods described and employed in the calculation of four-dimensional curvature in Chapter 14. Of all measures of the intrinsic curvature, one of the simplest is the Riemann scalar curvature invariant ${}^{(3)}R$ (written for simplicity of notation in what follows without the prefix, as R); and of all ways to define this invariant (see Chapter 14), one of the most compact uses the limit (see exercise 21.4)

Scalar curvature as measure of area deficit

$$R\begin{pmatrix}\text{at point}\\ \text{under study}\end{pmatrix} = \lim_{\varepsilon\to 0} 18\, \frac{4\pi\varepsilon^2 - \begin{pmatrix}\text{proper area of a surface (approximately}\\ \text{a 2-sphere) defined as the locus of the}\\ \text{points at a proper distance } \varepsilon\end{pmatrix}}{4\pi\,\varepsilon^4} \tag{21.50}$$

For a more detailed description of the curvature intrinsic to the 3-geometry, capitalize on differential geometry as already developed in Chapters 8 through 14, amending it only as required to distinguish what is three-dimensional from what is four-dimensional. Begin by considering a displacement

$$\boldsymbol{d\mathscr{P}} = \boldsymbol{e}_i\, \boldsymbol{d}x^i \tag{21.51}$$

within the hypersurface. Here the $\boldsymbol{e}_i$ are the basis tangent vectors $\boldsymbol{e}_i = \partial/\partial x^i$ (in one notation) or $\boldsymbol{e}_i = \partial\mathscr{P}/\partial x^i$ (in another notation) dual to the three coordinate 1-forms $\boldsymbol{d}x^i$. Any field of tangent vectors $\boldsymbol{A}$ that happens to lie in the hypersurface lets itself be expressed in terms of the same basis vectors:

$$\boldsymbol{A} = \boldsymbol{e}_i A^i. \tag{21.52}$$

The scalar product of this vector with the base vector $\boldsymbol{e}_j$ is

$$(\boldsymbol{A}\cdot\boldsymbol{e}_j) = A^i(\boldsymbol{e}_i\cdot\boldsymbol{e}_j) = A^i g_{ij} = A_j. \tag{21.53}$$

Now turn attention from a vector at one point to the parallel transport of the vector to a nearby point.

A vector lying on the equator of the Earth and pointing toward the North Star, transported parallel to itself along a meridian to a point still on the Earth's surface, but 1,000 km to the north, will no longer lie in the 2-geometry of the surface of the Earth. A telescope located in the northern hemisphere has to raise its tube to see the North Star! The generalization to a three-dimensional hypersurface imbedded in a 4-geometry is immediate. Take vector $\boldsymbol{A}$, lying in the hypersurface, and transport it along an elementary route lying in the hypersurface, and in the course of this transport displace it at each stage parallel to itself, where "parallel" means parallel with respect to the geometry of the enveloping 4-manifold. Then $\boldsymbol{A}$ will ordinarily end up no longer lying in the hypersurface. Thus the "covariant derivative" of $\boldsymbol{A}$ in the direction of the i-th coordinate direction in the geometry of the enveloping spacetime (that is, the $\boldsymbol{A}$ at the new point diminished by the transported $\boldsymbol{A}$) has the form (see §10.4)

$$^{(4)}\boldsymbol{\nabla}_{\boldsymbol{e}_i}\boldsymbol{A} = {}^{(4)}\boldsymbol{\nabla}_i\boldsymbol{A} = {}^{(4)}\boldsymbol{\nabla}_i(\boldsymbol{e}_j A^j) = \boldsymbol{e}_j\frac{\partial A^j}{\partial x^i} + ({}^{(4)}\Gamma^{\mu}_{\ ji}\boldsymbol{e}_\mu)A^j. \tag{21.54}$$

A special instance of this formula is the equation for the covariantly measured change of the base vector $\boldsymbol{e}_m$ itself,

$$^{(4)}\boldsymbol{\nabla}_i\boldsymbol{e}_m = {}^{(4)}\Gamma^{\mu}_{\ mi}\boldsymbol{e}_\mu. \tag{21.55}$$

In both (21.54) and (21.55) the presence of the "out-of-the-hypersurface component"

$$(A^{j(4)}\Gamma^{0}_{\ ji})(\boldsymbol{e}_0\cdot\boldsymbol{n}) \tag{21.56}$$

From parallel transport in 4-geometry to parallel transport in 3-geometry

is quite evident. Now kill this component. Project $^{(4)}\boldsymbol{\nabla}\boldsymbol{A}$ orthogonally onto the hypersurface. In this way arrive at a parallel transport and a covariant derivative that are intrinsic to the 3-geometry of the hypersurface. By rights this covariant derivative should be written $^{(3)}\boldsymbol{\nabla}$; but for simplicity of notation it will be written as $\boldsymbol{\nabla}$ in the rest of this chapter, except where ambiguity might arise. To get the value of the new covariant derivative, one has only to rewrite (21.54) with the suffix $^{(4)}$ replaced everywhere by a $^{(3)}$, or, better, dropped altogether and with the "dummy index" of summation $\mu = (0, 1, 2, 3)$ replaced by $m = (1, 2, 3)$. However, it is more convenient, following Israel (1966), to turn from an expression dealing with contravariant components A^i of $\boldsymbol{A}$ to one dealing with covariant components $A_i = (\boldsymbol{A}\cdot\boldsymbol{e}_i)$. Thus the covariant derivative of $\boldsymbol{A}$ in the direction of the i-th coordinate direction in the hypersurface, calculated with respect to the 3-geometry intrinsic to the hypersurface itself, has for its h-th covariant component the quantity [see equation (10.18)]

A new covariant derivative, taken with respect to the 3-geometry

$$A_{h|i} = \boldsymbol{e}_h\cdot{}^{(3)}\boldsymbol{\nabla}_{\boldsymbol{e}_i}\boldsymbol{A} \equiv \boldsymbol{e}_h\cdot\boldsymbol{\nabla}_i\boldsymbol{A} = \frac{\partial A_h}{\partial x^i} - A^m\Gamma_{mhi}\ (= A_{h;i} \text{ for } \boldsymbol{A} \text{ in } \Sigma). \tag{21.57}$$

Here the notation of the vertical stroke distinguishes this covariant derivative from the covariant derivative taken with respect to the 4-geometry, as, for example, in equations (10.17ff). The connection coefficients here for three dimensions, like those dealt with earlier for four dimensions [see the equations leading from (14.14) through (14.15)], allow themselves to be expressed in terms of the metric coefficients and their first derivatives, and have the interpretation

$$^{(3)}\Gamma_{mhi} \equiv \Gamma_{mhi} = \boldsymbol{e}_m\cdot\boldsymbol{\nabla}_i\boldsymbol{e}_h. \tag{21.58}$$

From the connection coefficients in turn, one calculates as in Chapter 14 the full Riemann curvature tensor $^{(3)}R^i_{\ jmn}$ of the 3-geometry intrinsic to the hypersurface.

Over and above the curvature intrinsic to the simultaneity, one now encounters a concept not covered in previous chapters (except fleetingly in Box 14.1), the *extrinsic curvature* of the 3-geometry. This idea has no meaning for a 3-geometry

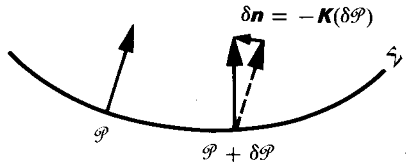


Figure 21.3.
Extrinsic curvature measures the fractional shrinkage and deformation of a figure lying in the spacelike hypersurface Σ that takes place when each point in the figure is carried forward a unit interval of proper time "normal" to the hypersurface out into the enveloping spacetime. (No enveloping spacetime? No extrinsic curvature!) The extrinsic curvature tensor is a positive multiple of the unit tensor when elementary displacements $\delta\mathscr{P}$, in whatever direction within the surface they point, all experience the same fractional shrinkage. Thus the extrinsic curvature of the hypersurface illustrated in the figure is positive. The dashed arrow represents the normal vector $\boldsymbol{n}$ at the fiducial point $\mathscr{P}$ after parallel transport to the nearby point $\mathscr{P} + \delta\mathscr{P}$.

conceived in and by itself. It depends for its existence on this 3-geometry's being imbedded as a well-defined slice in a well-defined enveloping spacetime. It measures the curvature of this slice relative to that enveloping 4-geometry (Figure 21.3).

Take the normal that now stands at the point $\mathscr{P}$ and, "keeping its base in the hypersurface" Σ, transport it parallel to itself as a "fiducial vector" to the point $\mathscr{P} + \delta\mathscr{P}$, and there subtract it from the normal vector that already stands at that point. The difference, $\delta\boldsymbol{n}$, may be regarded in the appropriate approximation as a "vector," the value of which is governed by and depends linearly on the "vector" of displacement $\delta\mathscr{P}$.

To obviate any appeal to the notion of approximation, go from the finite displacement $\delta\mathscr{P}$ to the limiting concept of the vector-valued "displacement 1-form" $\boldsymbol{d}\mathscr{P}$ [see equation 15.13]. Also replace the finite but not rigorously defined vector $\delta\boldsymbol{n}$ by the limiting concept of a vector-valued 1-form $\boldsymbol{dn}$. This quantity, regarded as a vector, being the change in a vector $\boldsymbol{n}$ that does not change in length, must represent a change in direction and thus stand perpendicular to $\boldsymbol{n}$. Therefore it can be regarded as lying in the hypersurface Σ. Depending linearly on $\boldsymbol{d}\mathscr{P}$, it can be represented in the form

Extrinsic curvature as an operator

$$\boldsymbol{dn} = -\boldsymbol{K}(\boldsymbol{d}\mathscr{P}). \tag{21.59}$$

Here the linear operator $\boldsymbol{K}$ is the extrinsic curvature presented as an abstract coordinate-independent geometric object. The sign of $\boldsymbol{K}$ as defined here is positive when the tips of the normals in Figure 21.3 are closer than their bases, as they are, for example, during the recontraction of a model universe, in agreement with the conventions employed by Eisenhart (1926), Schouten (1954), and Arnowitt, Deser and Misner (1962), but opposite to the convention of Israel (1966).

Into the slots in the 1-forms that appear on the lefthand and righthand sides of (21.59), insert in place of the general tangent vector [which is to describe the general local displacement, so far left open, as in the discussion following (2.12a)] a very special tangent vector, the basis vector $\boldsymbol{e}_i$, for a displacement in the i-th coordinate direction. Thus find (21.59) reading

$$^{(4)}\nabla_i \boldsymbol{n} = -\boldsymbol{K}(\boldsymbol{e}_i) = -K_i{}^j \boldsymbol{e}_j, \tag{21.60}$$

where the $K_i{}^j$ are the components of the linear operator $\boldsymbol{K}$ in a coordinate representation. Take the scalar product of both sides of (21.60) with the basis vector $\boldsymbol{e}_m$. Recall $(\boldsymbol{e}_m \cdot \boldsymbol{n}) = 0$. Thus establish the symmetry of the tensor K_{im}, covariantly presented, in its two indices:

$$\begin{aligned} K_{im} &= K_i{}^j g_{jm} = K_i{}^j(\boldsymbol{e}_j \cdot \boldsymbol{e}_m) = -\boldsymbol{e}_m \cdot {}^{(4)}\nabla_i \boldsymbol{n} = \boldsymbol{n} \cdot {}^{(4)}\nabla_i \boldsymbol{e}_m \\ &\underset{\uparrow}{=} (\boldsymbol{n} \cdot \boldsymbol{e}_0)^{(4)}\Gamma^0{}_{mi} = \boldsymbol{n} \cdot {}^{(4)}\nabla_m \boldsymbol{e}_i = K_{mi}. \end{aligned} \tag{21.61}$$

[see (21.55)]

A knowledge of the tensor K_{ij} of extrinsic curvature assists in revealing the changes of the four vectors $\boldsymbol{n}$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ under parallel transport. Equation (21.60) already tells how $\boldsymbol{n}$ changes under parallel transport. The change of $\boldsymbol{e}_m$ is to be read off from (21.55) as a vector. It is adequate identification of this vector to know its scalar product with each of four independent vectors: with the basis vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$, or, more briefly, with $\boldsymbol{e}_s$, in (21.58); and with the normal vector $\boldsymbol{n}$ in (21.61). Thus one arrives, following Israel (1966), at what are known as the equations of Gauss and Weingarten, in happy oversight of all change of notation in the intervening century:

Gauss-Weingarten equation for 4-transport in terms of extrinsic curvature

$$^{(4)}\nabla_i \boldsymbol{e}_j = K_{ij}\frac{\boldsymbol{n}}{\boldsymbol{n}\cdot\boldsymbol{n}} + {}^{(3)}\Gamma^h{}_{ji}\boldsymbol{e}_h. \tag{21.62}$$

Knowing from this equation how each basis vector in Σ changes, one also knows how to rewrite (21.54) for the change in any vector field $\boldsymbol{A}$ that lies in Σ. The change in both cases is expressed relative to a fiducial vector transported from a fiducial nearby point. By the term "parallel transport" one now means "parallel with respect to the geometry of the enveloping spacetime":

$$^{(4)}\nabla_i \boldsymbol{A} = A^j{}_{|i}\boldsymbol{e}_j + K_{ij}A^j\frac{\boldsymbol{n}}{(\boldsymbol{n}\cdot\boldsymbol{n})}. \tag{21.63}$$

Of special importance is the evaluation of extrinsic curvature when spacetime is sliced up into spacelike slices according to the plan of Arnowitt, Deser, and Misner as described in §21.4. The 4-geometry of the thin sandwich illustrated in Figure 21.2, rudimentary though it is, is fully defined by the 3-metric on the two faces of the sandwich and by the lapse and shift functions N and N^i. The normal in covariant representation according to (21.47) has the components

$$(n_0, n_1, n_2, n_3) = (-N, 0, 0, 0). \tag{21.64}$$

The change in $\boldsymbol{n}$ relative to "$\boldsymbol{n}$ transported parallel to itself in the enveloping 4-geometry," according to the definition of parallel transport, is

$$
\begin{aligned}
(\boldsymbol{dn})_i &= n_{i;k}\,\boldsymbol{dx}^k \\
&= \left[\frac{\partial n_i}{\partial x^k} - {}^{(4)}\Gamma^{\sigma}_{ik} n_\sigma\right]\boldsymbol{dx}^k \\
&= N\,{}^{(4)}\Gamma^{0}_{ik}\,\boldsymbol{dx}^k
\end{aligned}
\tag{21.65}
$$

Compare to the same change as expressed in terms of the extrinsic curvature tensor,

$$
(\boldsymbol{dn})_i = -K_{ik}\,\boldsymbol{dx}^k. \tag{21.66}
$$

Conclude that this tensor has the value

$$
K_{ik} = -n_{i;k} = -N\,{}^{(4)}\Gamma^{0}_{ik} = -N[{}^{(4)}g^{00}\,{}^{(4)}\Gamma_{0ik} + {}^{(4)}g^{0p}\,{}^{(4)}\Gamma_{pik}],
$$

or, with the help of equations (21.42) and (21.44),

Extrinsic curvature in terms of shift and change of 3-metric

$$
\begin{aligned}
K_{ik} &= (1/N)[{}^{(4)}\Gamma_{0ik} - N^{p}\,{}^{(3)}\Gamma_{pik}] \\
&= \frac{1}{2N}\left[\frac{\partial N_i}{\partial x^k} + \frac{\partial N_k}{\partial x^i} - \frac{\partial g_{ik}}{\partial t} - 2\Gamma_{pik}N^p\right] \\
&= \frac{1}{2N}\left[N_{i|k} + N_{k|i} - \frac{\partial g_{ik}}{\partial t}\right].
\end{aligned}
\tag{21.67}
$$

This is the extrinsic curvature expressed in terms of the ADM lapse and shift functions [Arnowitt, Deser, and Misner (1962)].

Extrinsic curvature of expanding 3-sphere

As an example, let Σ have the geometry of a 3-sphere

$$
ds^2 = a^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{21.68}
$$

Let the nearby spacelike slice in the one-parameter family of slices, the slice with the label $t + dt$ (only a label!) have a 3-metric given by the same formula with the radius a replaced by $a + da$. The 4-geometry of the thin sandwich between these two slices is completely undetermined until one gives the lapse and shift functions. For simplicity, take the shift vector N^i (see Figure 21.2) to be everywhere zero and the lapse function at every point on Σ to have the same value N. The separation in proper time between the two spheres is thus $d\tau = N\,dt$. Any geometric figure located in Σ expands with time. The fractional increase of any length in this figure per unit of proper time is the same in whatever direction that length is oriented, and has the value

$$
\begin{pmatrix}\text{fractional increase}\\ \text{of length per unit}\\ \text{of proper time}\end{pmatrix} = \frac{1}{a}\frac{da}{d\tau} = \frac{1}{2N}\frac{1}{a^2}\frac{d(a^2)}{dt}. \tag{21.69}
$$

The negative of this quantity, multiplied by the $\binom{1}{1}$ unit tensor, $\boldsymbol{1} = \boldsymbol{d\mathscr{P}}$, gives the extrinsic curvature tensor in $\binom{1}{1}$ representation,

$$
\boldsymbol{K} = -\frac{1}{2N}\frac{1}{a^2}\frac{d(a^2)}{dt}\boldsymbol{1}. \tag{21.70}
$$

One confirms this result (exercise 21.5) by direct calculation of the components K^i_i using the ADM formula (21.67) as the starting point.

The Riemann curvature $R^a{}_{bcd} = {}^{(3)}R^a{}_{bcd}$ intrinsic to the hypersurface Σ, together with the extrinsic curvature K_{ij}, give one information on the Riemann and Einstein curvatures of the 4-geometry. In the calculation, it is not convenient to use the coordinate basis,

$$\begin{array}{cc} \text{basis vectors,} & \text{basis 1-forms} \\ \boldsymbol{e}_0 = \partial_t, & \boldsymbol{d}t, \\ \boldsymbol{e}_i = \partial_i, & \boldsymbol{d}x^i, \end{array}$$

because ordinarily the basis vector $\boldsymbol{e}_0$ does not stand perpendicular to the hypersurface (see Figure 21.2). Adopt a different basis but one that is still self-dual,

Basic forms for calculating 4-curvature

$$\begin{array}{cc} \text{basis vectors,} & \text{basis 1-forms,} \\ \boldsymbol{e}_n \equiv \boldsymbol{n} = N^{-1}(\partial_t - N^m \partial_m), & \boldsymbol{\omega}^n = N\,\boldsymbol{d}t = (\boldsymbol{n}\cdot\boldsymbol{n})\boldsymbol{n} \\ \boldsymbol{e}_i = \partial_i, & \boldsymbol{\omega}^i \equiv \boldsymbol{d}x^i + N^i\,\boldsymbol{d}t. \end{array} \tag{21.71}$$

Also use Greek labels $\bar{\alpha} = n, 1, 2, 3$, instead of Greek labels $\alpha = 0, 1, 2, 3$, to list components.

Recall that curvature is measured by the change in a vector on transport around a closed route; or, from equation (14.23),

$$\mathcal{R}(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{w} = \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{w} - \boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{w} - \boldsymbol{\nabla}_{[\boldsymbol{u},\boldsymbol{v}]} \boldsymbol{w}. \tag{21.72}$$

Let the vector transported be $\boldsymbol{e}_i$ and let the route be defined by $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$. The latter two vectors belong to a coordinate basis. Therefore the "route closes automatically", $[\boldsymbol{e}_j, \boldsymbol{e}_k] = 0$, and the final term in (21.72) drops out of consideration. Call on (21.62) and (21.60) to find

$$\begin{aligned} {}^{(4)}\boldsymbol{\nabla}_{\boldsymbol{e}_j} {}^{(4)}\boldsymbol{\nabla}_{\boldsymbol{e}_k} \boldsymbol{e}_i &= {}^{(4)}\boldsymbol{\nabla}_{\boldsymbol{e}_j} \left[K_{ik} \frac{\boldsymbol{n}}{(\boldsymbol{n}\cdot\boldsymbol{n})} + {}^{(3)}\Gamma^m_{ik} \boldsymbol{e}_m \right] \\ &= K_{ik,j} \frac{\boldsymbol{n}}{(\boldsymbol{n}\cdot\boldsymbol{n})} - K_{ik} K_j{}^m \boldsymbol{e}_m \frac{1}{(\boldsymbol{n}\cdot\boldsymbol{n})} + {}^{(3)}\Gamma^m_{ik,j} \boldsymbol{e}_m \\ &\quad + {}^{(3)}\Gamma^m_{ik} \left[K_{mj} \frac{\boldsymbol{n}}{(\boldsymbol{n}\cdot\boldsymbol{n})} + {}^{(3)}\Gamma^s_{mj} \boldsymbol{e}_s \right]. \end{aligned} \tag{21.73}$$

Evaluate similarly the term with indices j and k reversed, subtract from (21.73), simplify, and find

Gauss-Codazzi: 4-curvature in terms of intrinsic 3-geometry and extrinsic curvature

$$\begin{aligned} \mathcal{R}(\boldsymbol{e}_j, \boldsymbol{e}_k)\boldsymbol{e}_i &= (K_{ik|j} - K_{ij|k}) \frac{\boldsymbol{n}}{(\boldsymbol{n}\cdot\boldsymbol{n})} \\ &\quad + [(\boldsymbol{n}\cdot\boldsymbol{n})^{-1}(K_{ij}K_k{}^m - K_{ik}K_j{}^m) + {}^{(3)}R^m{}_{ijk}]\boldsymbol{e}_m. \end{aligned} \tag{21.74}$$

The coefficients give directly the desired components of the curvature tensor

$${}^{(4)}R^m{}_{ijk} = {}^{(3)}R^m{}_{ijk} + (\boldsymbol{n}\cdot\boldsymbol{n})^{-1}(K_{ij}K_k{}^m - K_{ik}K_j{}^m) \tag{21.75}$$

and

$${}^{(4)}R^n{}_{ijk} = (\boldsymbol{n}\cdot\boldsymbol{n})^{-1\,(4)}R_{nijk} = -(\boldsymbol{n}\cdot\boldsymbol{n})^{-1}(K_{ij|k} - K_{ik|j}). \tag{21.76}$$

Equations (21.75) and (21.76) are known as the equations of Gauss and Codazzi [for literature, see Eisenhart (1926)]. It follows from (21.75) that the components of the curvature of the 3-geometry will normally only then agree with the corresponding components of the curvature of the 4-geometry when the imbedding happens to be accomplished at the point under study with a hypersurface free of extrinsic curvature. The directly opposite situation is illustrated by the familiar example of a 2-sphere imbedded in a flat 3-space, where the lefthand side of (21.75) (with dimensions lowered by one unit throughout!) is zero, and the extrinsic and intrinsic curvature on the right exactly cancel.

Important components of the Einstein curvature let themselves be evaluated from the Gauss-Codazzi results. In doing the calculation, it is simplest to think of $\boldsymbol{e}_i$, $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$ as being an orthonormal tetrad, $\boldsymbol{n}$ being itself already normalized and orthogonal to every vector in the hypersurface. Then, employing (14.7) and (21.75), one finds

Einstein curvature in terms of extrinsic curvature

$$\begin{aligned} -G^0_{\ 0} &= {}^{(4)}R^{12}{}_{12} + {}^{(4)}R^{23}{}_{23} + {}^{(4)}R^{31}{}_{31} \\ &= {}^{(3)}R^{12}{}_{12} + {}^{(3)}R^{23}{}_{23} + {}^{(3)}R^{31}{}_{31} \\ &\quad + (\boldsymbol{n}\cdot\boldsymbol{n})^{-1}[(K^2_1K^1_2 - K^2_2K^1_1) + (K^3_2K^2_3 - K^3_3K^2_2) \\ &\quad + (K^1_3K^3_1 - K^1_1K^3_3)] \\ &= \frac{1}{2}R - \frac{1}{2}(\boldsymbol{n}\cdot\boldsymbol{n})^{-1}[(\operatorname{Tr}\boldsymbol{K})^2 - \operatorname{Tr}(\boldsymbol{K}^2)]. \end{aligned} \tag{21.77}$$

Here R is the 3-dimensional scalar curvature invariant and Tr stands for "trace of"; thus,

$$\operatorname{Tr}\boldsymbol{K} = g^{ij}K_{ij} = g_{ij}K^{ij} = K^j_j \tag{21.78}$$

and

$$\operatorname{Tr}\boldsymbol{K}^2 = (K^2)^j_j = K_j^{\ m}K_m^{\ j} = g_{js}K^{sm}g_{mi}K^{ij}. \tag{21.79}$$

The result, though obtained in an orthonormal tetrad, plainly is covariant with respect to general coordinate transformations within the spacelike hypersurface; and it makes no explicit reference whatever to any time coordinate, in this respect providing a coordinate-free description of the Einstein curvature.

Equation (21.77) is the central Einstein equation, "mass-energy fixes curvature"

The Einstein field equation equates (21.77) to $8\pi\rho$, where ρ is the density of mass-energy. Expression (21.77) is the "measure of curvature that is independent of how curved one cuts a spacelike slice." This measure of curvature is central to the derivation of Einstein's field equation that is sketched in Box 17.2, item 3, "Physics on a Spacelike Slice."

The other component of the Einstein curvature tensor that is easily evaluated by (14.7) from the results at hand has the form

$$\begin{aligned} G^n_{\ 1} &= {}^{(4)}R^{n2}{}_{12} + {}^{(4)}R^{n3}{}_{13} \\ &= -(\boldsymbol{n}\cdot\boldsymbol{n})^{-1}(K^2_{1|2} - K^2_{2|1} + K^3_{1|3} - K^3_{3|1}), \end{aligned} \tag{21.80}$$

when referred to an orthonormal frame. One immediately translates to a form valid for any frame $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ in the hypersurface, orthonormal or not,

The other initial-value equation

$$G^n_{\ i} = -(\boldsymbol{n}\cdot\boldsymbol{n})^{-1}[K^m_{i|m} - (\operatorname{Tr}\boldsymbol{K})_{|i}]. \tag{21.81}$$

The Einstein field equation equates this quantity to 8π times the i-th covariant component of the density of momentum carried by matter and fields other than gravity.

The four components of the Einstein field equation so far written down will have a central place in what follows as "initial-value equations" of general relativity. The other six components will not be written out: (1) the dynamics lets itself be analyzed more simply by Hamiltonian methods; and (2) the calculation takes work. It demands that one evaluate the remaining type of object, $\mathcal{R}(\boldsymbol{e}_j, \boldsymbol{n})\boldsymbol{e}_i$. One step towards that calculation will be found in exercise 21.7. Sachs does the calculation (1964, equation 10) but only after specializing to Gaussian normal coordinates. These coordinates presuppose a very special slicing of spacetime: (1) geodesics issuing normally from the spacelike hypersurface $n = 0$ cut all subsequent simultaneities $n =$ constant normally; and (2) the n coordinate directly measures lapse of proper time, or proper length, whichever is appropriate,* along these geodesics. In coordinates so special it is not surprising that the answer looks simple:

$$^{(4)}R^{n}{}_{ink} = (\boldsymbol{n}\cdot\boldsymbol{n})^{-1}\left(\frac{\partial K_{ik}}{\partial n} + K_{im}K^{m}{}_{k}\right). \qquad \begin{pmatrix}\text{Gaussian normal}\\ \text{coordinates}\end{pmatrix} \tag{21.82}$$

Additional terms come into (21.82) when one uses, instead of the Gaussian normal coordinate system, the coordinate system of Arnowitt, Deser, and Misner. The ADM coordinates are employed here because they allow one to analyze the dynamics as one *wants* to analyze the dynamics, with freedom to push the spacelike hypersurface ahead in time at different rates in different places ("many-fingered time"). Fischer (1971) shows how to evaluate and understand the geometric content of such formulas in a coordinate-free way by using the concept of Lie derivative of a tensor field, an introduction to which is provided by exercise 21.8.

*Here Sachs' equation (10) is generalized to the case where the unit normal $\boldsymbol{n}$ is not necessarily timelike. Sachs used $\boldsymbol{n} = \partial/\partial t$.

EXERCISES

Exercise 21.4. SCALAR CURVATURE INVARIANT IN TERMS OF AREA DEFICIT

It being 10,000 km from North Pole to equator, one would have 62,832 km for the length of the "equator" if the earth were flat, as contrasted to the actual ~40,000 km, a difference reflecting the fact that the surface is curved up into closure. Turn from this "pre-problem" to the actual problem, a 3-sphere

$$ds^2 = a^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)].$$

Measure off from $\chi = 0$ a 2-sphere of proper radius $\varepsilon = a\chi$. Determine the proper area of this 2-sphere as a function of χ. Verify that relation (21.50) on the area deficit gives in the limit $\varepsilon \longrightarrow 0$ the correct result $R = 6/a^2$. For a more ambitious exercise: (1) take a general (smooth) 3-geometry; (2) express the metric near any chosen point in terms of Riemann's normal coordinates as given in §11.6; (3) determine the locus of the set of points at the proper distance ε to the lowest interesting power of ε in terms of the spherical polar angles θ and ϕ (direction of start of geodesic of length ε); (4) determine to the lowest interesting power of ε the proper area of the figure defined by these points; and thereby establish (21.50) [for more on this topic see, for example, Cartan (1946), pp. 252–256].

Exercise 21.5. EXTRINSIC CURVATURE TENSOR FOR SLICE OF FRIEDMANN GEOMETRY

Confirm the result (21.70) for the extrinsic curvature by direct calculation from formula (21.67).

Exercise 21.6. EVALUATION OF $\mathcal{R}(\boldsymbol{e}_j, \boldsymbol{e}_k)\boldsymbol{n}$

Evaluate this quantity along the model of (21.74) or otherwise. How can it be foreseen that the coefficient of $\boldsymbol{n}$ in the result must vanish identically? Comparing coefficients of $\boldsymbol{e}_m$, find ${}^{(4)}R^m{}_{njk}$ and test for equivalence to equation (21.76).

Exercise 21.7. EVALUATION OF THE COMMUTATOR $[\boldsymbol{e}_j, \boldsymbol{n}]$

The evaluation of this commutator is a first step toward the calculation of a quantity like $\mathcal{R}(\boldsymbol{e}_j, \boldsymbol{n})\boldsymbol{e}_i$. Expressing $\boldsymbol{e}_j$ as the differential operator $\partial/\partial x^j$, use (21.49) to represent $\boldsymbol{n}$ also as a differential operator. In this way, show that the commutator in question has the value $-(N_{,j}/N)\boldsymbol{n} - (N^m{}_{,j}/N)\boldsymbol{e}_m$.

Exercise 21.8. LIE DERIVATIVE OF A TENSOR (exercise provided by J. W. York, Jr.)

Define the Lie derivative of a tensor field and explore some of its properties. The Lie derivative along a vector field $\boldsymbol{n}$ is a differential operator that operates on tensor fields $\boldsymbol{T}$ of type $\binom{r}{s}$, converting them into tensors $\pounds_{\boldsymbol{n}}\boldsymbol{T}$, also of type $\binom{r}{s}$. The Lie differentiation process obeys the usual chain rule and has additivity properties [compare equations (10.2b, 10.2c, 10.2d) for the covariant derivative]. For scalar functions f, one has $\pounds_{\boldsymbol{n}} f \equiv \boldsymbol{n}[f] = f_{,\mu}n^\mu$. The Lie derivative of a vector field $\boldsymbol{u}$ along a vector field $\boldsymbol{v}$ was defined in exercise 9.11 by

$$\pounds_{\boldsymbol{u}}\boldsymbol{v} \equiv [\boldsymbol{u}, \boldsymbol{v}].$$

If the action of $\pounds_{\boldsymbol{n}}$ on 1-forms is defined, the extension to tensors of general type will be simple, because the latter can always be decomposed into a sum of tensor products of vectors and 1-forms. If $\boldsymbol{\sigma}$ is a 1-form and $\boldsymbol{v}$ is a vector, then one defines $\pounds_{\boldsymbol{n}}\boldsymbol{\sigma}$ to be that 1-form satisfying

$$\langle \pounds_{\boldsymbol{n}}\boldsymbol{\sigma}, \boldsymbol{v}\rangle = \boldsymbol{n}[\langle \boldsymbol{\sigma}, \boldsymbol{v}\rangle] - \langle \boldsymbol{\sigma}, [\boldsymbol{n}, \boldsymbol{v}]\rangle$$

for arbitrary $\boldsymbol{v}$.

(a) Show that in a coordinate basis

$$\pounds_{\boldsymbol{n}}\boldsymbol{\sigma} = (\sigma_{\alpha,\beta}n^\beta + \sigma_\beta n^\beta{}_{,\alpha})\,\boldsymbol{d}x^\alpha.$$

(b) Show that in a coordinate basis

$$\pounds_{\boldsymbol{n}}\boldsymbol{T} = (T_{\alpha\beta,\mu}n^\mu + T_{\mu\beta}n^\mu{}_{,\alpha} + T_{\alpha\mu}n^\mu{}_{,\beta})\,\boldsymbol{d}x^\alpha \otimes \boldsymbol{d}x^\beta$$

where $\boldsymbol{T}$ is of type $\binom{0}{2}$.

(c) Show that in (a) and (b), all partial derivatives can be replaced by covariant derivatives. [Observe that Lie differentiation is defined independently of the existence of an affine connection. For more information, see, for example, Bishop and Goldberg (1968) and Schouten (1954)].

Exercise 21.9. EXPRESSION FOR DYNAMIC COMPONENTS OF THE CURVATURE TENSOR (exercise provided by J. W. York, Jr.)

The Gauss-Codazzi equations can be viewed as giving 14 of the 20 algebraically independent components of the spacetime curvature tensor in terms of the intrinsic and extrinsic geometry of three-dimensional (non-null) hypersurfaces. In order to accomplish a space-plus-time splitting of the Hilbert Lagrangian $\sqrt{-g}\,{}^{(4)}R$, one must express, in addition, the remaining

2

6 components of the curvature tensor in an analogous manner. It is convenient for this purpose to express all tensors as spacetime tensors, and to use Lie derivation in the direction of the timelike unit normal field of the spacelike hypersurfaces as a generalized notion of time differentiation. A number of preliminary results must be proven:

(a)
$$\mathcal{L}_{\boldsymbol{u}} g_{\mu\nu} = u_{\mu;\nu} + u_{\nu;\mu},$$

(b)
$$\begin{aligned}\mathcal{L}_{\boldsymbol{u}}(g_{\mu\nu} + u_\mu u_\nu) &\equiv \mathcal{L}_{\boldsymbol{u}}(\gamma_{\mu\nu})\\ &= u_{\mu;\nu} + u_{\nu;\mu} + u_\mu a_\nu + a_\mu u_\nu,\end{aligned}$$

where $\gamma_{\mu\nu}$ is the metric of the spacelike hypersurface, expressed in the spacetime coordinate basis, and $a^\mu \equiv u^\lambda \nabla_\lambda u^\mu$ is the curvature vector (4-acceleration) of the timelike normal curves whose tangent field is u^μ. (Recall that $u_\mu a^\mu = 0$.)

(c) Prove that the extrinsic curvature tensor is given by

$$K_{\mu\nu} = -\frac{1}{2} \mathcal{L}_{\boldsymbol{u}} \gamma_{\mu\nu}.$$

(d) The unit tensor of projection into the hypersurface is defined by

$$\perp^\mu_\nu \equiv \delta^\mu_\nu + u^\mu u_\nu.$$

In terms of $\perp$ show that one can write

$$u_{\alpha;\beta} \equiv -K_{\alpha\beta} - \omega_{\alpha\beta} - a_\alpha u_\beta,$$

where

$$K_{\alpha\beta} = -\perp^\mu_\alpha \perp^\nu_\beta u_{(\mu;\nu)}$$

and

$$\omega_{\alpha\beta} = -\perp^\mu_\alpha \perp^\nu_\beta u_{[\mu;\nu]}.$$

(e) From the fact that u^μ is the unit normal field for a family of spacelike hypersurfaces, show that $\omega_{\alpha\beta} = 0$.

(f) The needed tools are now on hand. To obtain the result:

(i) Write down $\mathcal{L}_{\boldsymbol{u}} K_{\mu\nu}$ (see exercise 21.8);

(ii) Insert this expression into the Ricci identity in the form

$$u^\sigma \nabla_\sigma \nabla_\mu u_\nu = u^\sigma \nabla_\mu \nabla_\sigma u_\nu + {}^{(4)}R_{\rho\nu\mu\sigma} u^\sigma u^\rho;$$

(iii) Project the two remaining free indices into the hypersurface using $\perp$, and show that one obtains

$$\begin{aligned}\perp^\mu_\alpha \perp^\rho_\beta {}^{(4)}R_{\mu\nu\rho\sigma} u^\nu u^\sigma &= \mathcal{L}_{\boldsymbol{u}} K_{\alpha\beta} + K_{\alpha\gamma} K^\gamma_\beta\\ &+ {}^{(3)}\nabla_{(\alpha} a_{\beta)} + a_\alpha a_\beta,\end{aligned}$$

where ${}^{(3)}\nabla_\alpha a_\beta \equiv \perp^\mu_\alpha \perp^\nu_\beta \nabla_\mu a_\nu$ can be shown to be the three-dimensional covariant derivative of a_β. In Gaussian normal coordinates, show that one obtains from this result

$$R_{0i0j} = \frac{\partial}{\partial t} K_{ij} + K_{ik} K^k_j.$$

(g) Finally, in the construction of ${}^{(4)}R$, one needs to show that

$$\gamma^{\mu\nu}[{}^{(3)}\nabla_{(\mu} a_{\nu)} + a_\mu a_\nu] = g^{\mu\nu}[{}^{(3)}\nabla_{(\mu} a_{\nu)} + a_\mu a_\nu] = a^\lambda_{;\lambda}.$$

Exercise 21.10. EXPRESSION OF ${}^{(4)}R^i{}_{nin}$ IN TERMS OF EXTRINSIC CURVATURE, PLUS A COVARIANT DIVERGENCE (exercise provided by K. Kuchař)

Let α' be an arbitrary smooth set of four coordinates, not necessarily coordinated in any way with the choice of the 1-parameter family of hypersurfaces.

(a) Show that

$${}^{(4)}R^i{}_{nin} = g^{\alpha'\gamma'}n^{\beta'}(n_{\alpha';\beta'\gamma'} - n_{\alpha';\gamma'\beta'}).$$

(b) Show that the covariant divergences

$$(n^{\beta'}n^{\gamma'}{}_{;\beta'})_{;\gamma'}$$

and

$$-(n^{\beta'}n^{\gamma'}{}_{;\gamma'})_{;\beta'}$$

can be removed from this expression in such a way that what is left behind contains only first derivatives of the unit normal vector $\boldsymbol{n}$.

(c) Noting that the basis vectors $\boldsymbol{e}_i$ and $\boldsymbol{n}$ form a complete set, justify the formula

$$g^{\beta'\mu'} = e_i{}^{\beta'}\omega^{i\mu'} + (\boldsymbol{n}\cdot\boldsymbol{n})^{-1}n^{\beta'}n^{\mu'},$$

where ω^i is the 1-form dual to $\boldsymbol{e}_i$.

(d) Noting that $n_{\alpha';\beta'}n^{\alpha'} = 0$ and

$$K_{ij} = -e_{i\alpha'}n^{\alpha'}{}_{;\beta'}e_j{}^{\beta'},$$

show that

$${}^{(4)}R^i{}_{nin} = (\mathrm{Tr}\boldsymbol{K})^2 - \mathrm{Tr}\boldsymbol{K}^2 \text{ plus a covariant divergence.}$$

---

§21.6. THE HILBERT ACTION PRINCIPLE AND THE ARNOWITT-DESER-MISNER MODIFICATION THEREOF IN THE SPACE-PLUS-TIME SPLIT

For analyzing the dynamics, it happily proves unnecessary to possess the missing formula for ${}^{(4)}R^n{}_{ink}$. It is essential, however, to have the Lagrangian density,

$$16\pi\mathcal{L}_{\mathrm{geom}} = (-{}^{(4)}g)^{1/2}\,{}^{(4)}R, \tag{21.83}$$

in the Hilbert action principle as the heart of all the dynamic analysis. In the present ADM (1962) notation, this density has the form

$$\begin{aligned}(-{}^{(4)}g)^{1/2}\,{}^{(4)}R &= (-{}^{(4)}g)^{1/2}[{}^{(4)}R^{ij}{}_{ij} + 2\,{}^{(4)}R^{in}{}_{in}]\\ &= (-{}^{(4)}g)^{1/2}[R + (\boldsymbol{n}\cdot\boldsymbol{n})(\mathrm{Tr}\,\boldsymbol{K}^2 - (\mathrm{Tr}\,\boldsymbol{K})^2) + 2(\boldsymbol{n}\cdot\boldsymbol{n})\,{}^{(4)}R^i{}_{nin}].\end{aligned} \tag{21.84}$$

Drop a complete derivative from the Hilbert action principle to get the ADM principle

Kuchař (1971b; see also exercise 21.10) shows how to calculate a sufficient part of this quantity without calculating all of it. The difference between the "sufficient part" and the "whole" is a time derivative plus a divergence, a quantity of the form

$$[(-{}^{(4)}g)^{1/2}A^\alpha]_{,\alpha} = (-{}^{(4)}g)^{1/2}A^\alpha{}_{;\alpha}. \tag{21.85}$$

2

When one multiplies (21.83) by $dt\, dx^1\, dx^2\, dx^3$ and integrates to obtain the action integral, the term (21.85) integrates out to a surface term. Variations of the geometry interior to this surface make no difference in the value of this surface term. Therefore it has no influence on the equations of motion to drop the term (21.85). The result of the calculation (exercise 21.10) is simple: what is left over after dropping the divergence merely changes the sign of the terms in $\mathrm{Tr}\,\boldsymbol{K}^2$ and $(\mathrm{Tr}\,\boldsymbol{K})^2$ in (21.84). Thus the variation principle becomes

$$\begin{aligned}(\text{extremum}) &= I_{\text{modified}} = \int \mathcal{L}_{\text{modified}}\, d^4x \\ &= (1/16\pi)\int [R + (\boldsymbol{n}\cdot\boldsymbol{n})((\mathrm{Tr}\,\boldsymbol{K})^2 - \mathrm{Tr}\,\boldsymbol{K}^2)]Ng^{1/2}\, dt\, d^3x + \int \mathcal{L}_{\text{fields}}\, d^4x.\end{aligned} \tag{21.86}$$

This expression, rephrased, is the starting point for Arnowitt, Deser, and Misner's analysis of the dynamics of geometry.

Two supplements from a paper of York (1972b; see also exercise 21.9) enlarge one's geometric insight into what is going on in the foregoing analysis. First, the tensor of extrinsic curvature lets itself be defined [see also Fischer (1971)] most naturally in the form

$$\boldsymbol{K} = -\frac{1}{2}\boldsymbol{\pounds}_{\boldsymbol{n}}\boldsymbol{g}, \tag{21.87}$$

where $\boldsymbol{g}$ is the metric tensor of the 3-geometry, $\boldsymbol{n}$ is the timelike unit normal field, and $\boldsymbol{\pounds}$ is the Lie derivative as defined in exercise 21.8. Second, the divergence (21.85), which has to be added to the Lagrangian of (21.86) to obtain the full Hilbert Lagrangian, is

$$-2[(-{}^{(4)}g)^{1/2}(n^{\alpha'}\mathrm{Tr}\,\boldsymbol{K} + a^{\alpha'})]_{,\alpha'}, \tag{21.88}$$

where the coordinates are general (see exercise 21.10), and

$$a^{\alpha'} = n^{\alpha'}{}_{;\beta'}n^{\beta'} \tag{21.89}$$

is the 4-acceleration of an observer traveling along the timelike normal $\boldsymbol{n}$ to the successive slices.

§21.7. THE ARNOWITT, DESER, AND MISNER FORMULATION OF THE DYNAMICS OF GEOMETRY

Dirac (1959, 1964, and earlier references cited therein) formulated the dynamics of geometry in a (3 + 1)-dimensional form, using generalizations of Poisson brackets and of Hamilton equations. Arnowitt, Deser, and Misner instead made the Hilbert-Palatini variational principle the foundation for this dynamics. Because of its simplicity, this ADM (1962) approach is followed here. The gravitational part of the integrand in the Hilbert-Palatini action principle is rewritten in the condensed but standard form (after inserting a 16π that ADM avoid by other units) as

$$\begin{aligned}16\pi\mathcal{L}_{\text{geom true}} = \mathcal{L}_{\text{geom ADM}} &= -g_{ij}\,\partial\pi^{ij}/\partial t - N\mathcal{H} - N_i\mathcal{H}^i \\ &\quad - 2\left[\pi^{ij}N_j - \frac{1}{2}N^i\,\mathrm{Tr}\,\boldsymbol{\pi} + N^{|i}(g)^{1/2}\right]_{,i}.\end{aligned} \tag{21.90}$$

Here each item of abbreviation has its special meaning and will play its special part, a part foreshadowed by the name now given it:

$$\pi^{ij}_{\text{true}} = \frac{\delta(\text{action})}{\delta g_{ij}} = \begin{pmatrix}\text{“geometrodynamic} \\ \text{field momentum” dyn-} \\ \text{amically conjugate to} \\ \text{the “geometrodynamic} \\ \text{field coordinate” } g_{ij}\end{pmatrix} = \frac{\pi^{ij}}{16\pi};\ \pi^{ij} = g^{1/2}(g^{ij}\operatorname{Tr}\boldsymbol{K} - K^{ij}) \tag{21.91}$$

Momenta conjugate to the dynamic g_{ij}

(here the π^{ij} of ADM is usually more convenient than π^{ij}_{true}); and

$$\begin{aligned}\mathcal{H}_{\text{true}} &= \mathcal{H}(\pi^{ij}_{\text{true}}, g_{ij}) = (\text{“super-Hamiltonian”}) = \mathcal{H}/16\pi; \\ \mathcal{H}(\pi^{ij}, g_{ij}) &= g^{-1/2}\left(\operatorname{Tr}\boldsymbol{\pi}^2 - \frac{1}{2}(\operatorname{Tr}\boldsymbol{\pi})^2\right) - g^{1/2}R;\end{aligned} \tag{21.92}$$

and

$$16\pi\mathcal{H}^i_{\text{true}} = \mathcal{H}^i = \mathcal{H}^i(\pi^{ij}, g_{ij}) = (\text{“supermomentum”}) = -2\pi^{ik}{}_{|k}. \tag{21.93}$$

Here the covariant derivative is formed treating π^{ik} as a tensor density, as its definition in (21.91) shows it to be (see §21.2). The quantities to be varied to extremize the action are the coefficients in the metric of the 4-geometry, as follows: the six g_{ij} and the lapse function N and shift function N_i; and also the six "geometrodynamic momenta," π^{ij}. To vary these momenta as well as the metric is (1) to follow the pattern of elementary Hamiltonian dynamics (Box 21.1), where, by taking the momentum p to be as independently variable as the coordinate x, one arrives at two Hamilton equations of the first order instead of one Lagrange equation of the second order, and (2) to follow in some measure the lead of the Palatini variation principle of §21.2. There, however, one had 40 connection coefficients to vary, whereas here one has come down to only six π^{ij}. To know these momenta and the 3-metric is to know the extrinsic curvature. Before carrying out the variation, drop the divergence $-2[\quad]_{,i}$ from (21.90), since it gives rise only to surface integrals and therefore in no way affects the equations of motion that will come out of the variational principle. Also rewrite the first term in (21.90) in the form

$$-(\partial/\partial t)(g_{ij}\pi^{ij}) + \pi^{ij}\,\partial g_{ij}/\partial t, \tag{21.94}$$

and drop the complete time-derivative from the variation principle, again because it is irrelevant to the resulting equations of motion. The action principle now takes the form

$$\begin{aligned}\text{extremum} &= I_{\text{true}} = I_{\text{ADM}}/16\pi \\ &= (1/16\pi)\int[\pi^{ij}\,\partial g_{ij}/\partial t - N\mathcal{H}(\pi^{ij}, g_{ij}) - N_i\mathcal{H}^i(\pi^{ij}, g_{ij})]\,d^4x \\ &\qquad + \int \mathcal{L}_{\text{field}}\,d^4x.\end{aligned} \tag{21.95}$$

The action principle itself, here as always, tells one what must be fixed to make the action take on a well-defined value (if and when the action possesses an extremum). Apart from appropriate potentials having to do with fields other than geom-

Action principle says, fix 3-geometry on each face of sandwich

etry, the only quantities that have to be fixed appear at first sight to be the values of the six g_{ij} on the initial and final spacelike hypersurfaces. However, the ADM action principle is invariant with respect to any change of coordinates x^1, x^2, x^3 $\longrightarrow x^{\bar{1}}$, $x^{\bar{2}}$, $x^{\bar{3}}$ within the successive spacelike slices. Therefore the quantities that really have to be fixed on the two faces of the sandwich are the 3-geometries ${}^{(3)}\mathcal{G}'$ (on the initial hypersurface) and ${}^{(3)}\mathcal{G}$ (on the final hypersurface) and nothing more.

What a 3-geometry is

In mathematical terms, a 3-geometry ${}^{(3)}\mathcal{G}$ is the "equivalence class" of a set of differentiable manifolds that are isometrically equivalent to each other under diffeomorphisms. In the terms of the everyday physicist, a 3-geometry is the equivalence class of 3-metrics $g_{ij}(x, y, z)$ that are equivalent to one another under coordinate transformations. In more homely terms, two automobile fenders have one and the same 2-geometry if they have the same shape, regardless of how much the coordinate rulings painted on the one may differ from the coordinate rulings painted on the other.

Electromagnetism gives example of momentum conjugate to "field coordinate"

To have in equation (21.95) an example of a field Lagrangian that is at the same time physically relevant and free of avoidable complications, take the case of a source-free electromagnetic field. It would be possible to take the field Lagrangian to have the standard Maxwell value,

$$(1/8\pi)(\boldsymbol{E}^2 - \boldsymbol{B}^2) \longrightarrow -(1/16\pi)F_{\mu\nu}F^{\mu\nu}, \tag{21.96}$$

with

$$F_{\mu\nu} = \partial A_\nu/\partial x^\mu - \partial A_\mu/\partial x^\nu. \tag{21.97}$$

The variation of the Lagrangian with respect to the independent dynamic variables of the field, the four potentials A_α, would then immediately give the four second-order partial differential wave equations for these four potentials. However, to have instead a larger number of first-order equations is as convenient for electrodynamics as it is for geometrodynamics. One seeks for the analog of the Hamiltonian equations of particle dynamics,

$$\begin{aligned} dx/dt &= \partial H(x,p)/\partial p, \\ dp/dt &= -\partial H(x,p)/\partial x. \end{aligned} \tag{21.98}$$

One gets those equations by replacing the Lagrange integral $\int L(x, \dot{x})\, dt$ by the Hamilton integral $\int [p\dot{x} - H(x,p)]\, dt$. Likewise, here one replaces the action integrand of (21.96) by what in flat spacetime would be

$$(1/4\pi)\left[A_{\mu,\nu}F^{\mu\nu} + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\right]. \tag{21.99}$$

In actuality, spacetime is to be regarded as not only curved but also sliced up into spacelike hypersurfaces. This (3 + 1) split of the geometry made it desirable to split the ten geometrodynamic potentials into the six g_{ij} and the four lapse and shift functions. Here one similarly splits the four A_μ into the three components A_i of the vector potential and the scalar potential $A_0 = -\phi$ (with the sign so chosen that, in flat spacetime in a Minkowski coordinate system, $\phi = A^0$). In this notation, the

Lagrange density function, including the standard density factor $(-{}^{(4)}g)^{1/2}$ but dropping a complete time integral $(\partial/\partial t)(A_i \mathcal{E}^i)$ that has no influence on the equations of motion, is given by the formula

Lagrange density for electromagnetism

$$\begin{aligned}4\pi\mathcal{L}_{\text{field}} = &-\mathcal{E}^i\, \partial A_i/\partial t + \phi \mathcal{E}^i{}_{,i}\\ &-\frac{1}{2}Ng^{-1/2}g_{ij}(\mathcal{E}^i\mathcal{E}^j + \mathcal{B}^i\mathcal{B}^j) + N^i[ijk]\mathcal{E}^j\mathcal{B}^k.\end{aligned} \tag{21.100}$$

Here use is made of the alternating symbol $[ijk]$, defined as changing sign on the interchange of any two labels, and normalized so that $[123] = 1$. Note that the 3-tensor ε^{ijk} and the alternating symbol $[ijk]$ are related much as are the corresponding four-dimensional objects in equation (8.10), so that one can write

$$\mathcal{B}^i = \frac{1}{2}[ijk](A_{k,j} - A_{j,k}). \tag{21.101}$$

The quantities $\mathcal{B}^i$ are the components of the magnetic field in the spacelike slice. They are not regarded as independently variable. They are treated as fully fixed by the choice of the three potentials A_i. The converse is the case for the components $\mathcal{E}^i$ of the electric field: they are treated like momenta, and as independently variable.

Extremizing the action with respect to the $\mathcal{E}^i$ (exercise 21.11) gives the analog of the equation $dx/dt = p/m$ in particle mechanics, and the analog of the equation

$$E_i = -\partial A_i/\partial t - \partial\phi/\partial x^i \tag{21.102}$$

of flat-spacetime electrodynamics; namely,

The initial-value equation of electromagnetism

$$-\partial A_i/\partial t - \phi_{,i} - Ng^{-1/2}g_{ij}\mathcal{E}^j - [ijk]N^j\mathcal{B}^k = 0. \tag{21.103}$$

Here the last term containing the shift functions N^j, arises from the obliquity of the coordinate system. ADM give the following additional but equivalent ways to state the result (21.103):

$$\begin{aligned}\mathcal{E}^i &= \frac{1}{2}[ijk]\,{}^*F_{jk}\\ &= \frac{1}{2}[ijk]\left\{\frac{1}{2}[jk\mu\nu](-{}^{(4)}g)^{1/2\,(4)}g^{\mu\alpha\,(4)}g^{\nu\beta}F_{\alpha\beta}\right\}.\end{aligned} \tag{21.104}$$

They note that $\mathcal{E}^j$ and $\mathcal{B}^j$ are not directly the contravariant components of the fields in the simultaneity Σ,

$$\boldsymbol{E} = E^j\boldsymbol{e}_j,\ \boldsymbol{B} = B^j\boldsymbol{e}_j, \tag{21.105}$$

but the contravariant densities,

$$\mathcal{E}^j = g^{1/2}E^j,\ \mathcal{B}^j = g^{1/2}B^j. \tag{21.106}$$

Extremizing the action with respect to the three A_i (exercise 21.12) gives the curved-spacetime analog of the Maxwell equations,

$$\partial\boldsymbol{E}/\partial t = \boldsymbol{\nabla}\times\boldsymbol{B}. \tag{21.107}$$

Divergence relation by extremization with respect to ϕ

The remaining potential, ϕ, enters the action principle at only one point. Extremizing with respect to it gives immediately the divergence relation of source-free electromagnetism,

$$\mathcal{E}^i{}_{,i} = 0. \tag{21.108}$$

Action principle tells what to fix at limits

If an action principle tells in and by itself what quantities are to be fixed at the limits, what lessons does (21.100) give on this score? One can go back to the example of particle mechanics in Hamiltonian form, as in Box 21.1, and note that there the momentum p could "flap in the breeze." Only the coordinate x had to be fixed at the limits. Thus the variation of the action was

$$\begin{aligned}\delta I &= \delta \int [p\dot{x} - H(x,p)]\, dt \qquad (21.109)\\ &= \int \{[\dot{x} - \partial H/\partial p]\, \delta p + (d/dt)(p\, \delta x) + [-\dot{p} - \partial H/\partial x]\, \delta x\}\, dt.\end{aligned}$$

To arrive at a well-defined extremum of the action integral I, it was not enough to annul the coefficients, in square brackets, of δp and δx; that is, to impose Hamilton's equations of motion. It was necessary in addition to annul the quantities at limits, $p\, \delta x$; that is, to specify x at the start and at the end of the motion. Similarly here. The quantities ϕ and $\mathcal{E}^i$ flap in the breeze, but the magnetic field has to be specified on the two faces of the sandwich to allow one to speak of a well-defined extremum of the action principle. Why the magnetic field, or the three quantities

At limits, fix not potentials but magnetic field itself

$$\partial A_j/\partial x^i - \partial A_i/\partial x^j; \tag{21.110}$$

why not the three A_i themselves? When one varies (21.100) with respect to the A_i, and integrates the variation of the first term by parts, as one must to arrive at the dynamic equations, one obtains a term at limits

$$\int_{\Sigma_{\text{initial}}} \mathcal{E}^i\, \delta A_i\, d^3x - \int_{\Sigma_{\text{final}}} \mathcal{E}^i\, \delta A_i\, d^3x. \tag{21.111}$$

One demands that both these terms at limits must vanish in order to have a well-defined variational problem. Go from the given vector potential to another vector potential, $A_{i_{\text{new}}}$, by the gauge transformation

$$A_{i_{\text{new}}} = A_i + \delta A_i = A_i + \partial\lambda/\partial x^i. \tag{21.112}$$

The magnetic-field components given by the three $A_{i_{\text{new}}}$ differ in no way from those listed in (21.110). Moreover the "variation at limits,"

$$\int \mathcal{E}^i\, \delta A_i\, d^3x = \int \mathcal{E}^i\, \partial\lambda/\partial x^i\, d^3x = -\int \lambda \mathcal{E}^i{}_{,i}\, d^3x, \tag{21.113}$$

is automatically zero by virtue of the divergence condition (21.108), for any arbitrary choice of λ. Therefore the quantities fixed at limits are not the three A_i themselves (mere potentials) but the physically significant quantities (21.110), the components of the magnetic field. Moreover, the divergence condition $\mathcal{E}^i{}_{,i} = 0$ now becomes *the initial-value equation for the determination of the potential* ϕ.

In the preceding paragraph one need only replace "the three A_i" by "the six g_{ij}" and "the components of the magnetic field" by "the 3-geometry ${}^{(3)}\mathcal{G}$" and "the potential ϕ" by "the lapse and shift functions N and N^i" to pass from electrodynamics to geometrodynamics.

With this parallelism in view, turn back to the variational principle (21.95) of general relativity in the ADM formulation. With the 3-geometry fixed on the two faces of the sandwich, vary conditions in between to extremize the action, varying in turn the π^{ij}, the g_{ij}, and the lapse and shift functions. The geometrodynamic momenta appear everywhere only algebraically in the action principle, except in the term $-2N_i\pi^{ij}{}_{|j}$. Variation and integration by parts gives $2N_{i|j}\,\delta\pi^{ij}$. Collecting coefficients of $\delta\pi^{ij}$ and annuling the sum of these coefficients, one arrives at one of the several conditions required for an extremum,

ADM principle reproduces formula for geometrodynamic momentum

$$\partial g_{ij}/\partial t = 2Ng^{-1/2}\left(\pi_{ij} - \frac{1}{2}g_{ij}\mathrm{Tr}\boldsymbol{\pi}\right) + N_{i|j} + N_{j|i}. \tag{21.114}$$

This result agrees with what one gets from equations (21.91) defining geometrodynamic momentum in terms of extrinsic curvature, together with expression (21.67) for extrinsic curvature in terms of lapse and shift. The result (21.114) here is no less useful than the result

$$dx/dt = \partial H(x,p)/\partial p = p/m$$

in the most elementary problem in mechanics: it marks the first step in splitting a second-order equation or equations into twice as many first-order equations.

Now vary the action with respect to the g_{ij} and again, after appropriate integration by parts and rearrangement, find the remaining first-order dynamic equations of general relativity [simplified by use of equations (21.116) and (21.117)],

Dynamic and initial-value equations out of ADM formalism

$$\begin{aligned}\partial\pi^{ij}/\partial t = {} & -Ng^{1/2}\left(R^{ij} - \frac{1}{2}g^{ij}R\right) + \frac{1}{2}Ng^{-1/2}g^{ij}\left(\mathrm{Tr}\,\boldsymbol{\pi}^2 - \frac{1}{2}(\mathrm{Tr}\,\pi)^2\right)\\ & - 2Ng^{-1/2}\left(\pi^{im}\pi_m{}^j - \frac{1}{2}\pi^{ij}\mathrm{Tr}\,\boldsymbol{\pi}\right)\\ & + g^{1/2}(N^{|ij} - g^{ij}N^{|m}{}_{|m}) + (\pi^{ij}N^m)_{|m} \\ & - N^i{}_{|m}\pi^{mj} - N^j{}_{|m}\pi^{mi} + \begin{bmatrix}\text{source terms arising from fields}\\ \text{other than geometry, omitted here for}\\ \text{simplicity, but discussed by ADM (1962)}\end{bmatrix}^{ij}.\end{aligned} \tag{21.115}$$

Finally extremize the action (21.95) with respect to the lapse function N and the shift functions N_i, and find the four so-called initial-value equations of general relativity, equivalent to (21.77) and (21.81) or to $G^\alpha_n = 8\pi T^\alpha_n$; thus,

$$-(1/16\pi)\mathcal{H}(\pi^{ij}, g_{ij}) = (1/8\pi)Ng^{-1/2}g_{ij}(\mathcal{E}^i\mathcal{E}^j + \mathcal{B}^i\mathcal{B}^j), \tag{21.116}$$

$$-(1/16\pi)\mathcal{H}^i(\pi^{ij}, g_{ij}) = -(1/4\pi)[ijk]\mathcal{E}^j\mathcal{B}^k. \tag{21.117}$$

EXERCISES

Exercise 21.11. FIRST EXPLOITATION OF THE ADM VARIATIONAL PRINCIPLE FOR THE ELECTROMAGNETIC FIELD

Extremize the action principle (21.100) with respect to the $\mathcal{E}^i$ and derive the result (21.103).

Exercise 21.12. SECOND EXPLOITATION OF THE ADM VARIATIONAL PRINCIPLE FOR THE ELECTROMAGNETIC FIELD

Extremize (21.100) with respect to the A_i, and verify that the resulting equations in any Minkowski-flat region are equivalent to (21.107).

Exercise 21.13. FARADAY-MAXWELL SOURCE TERM IN THE DYNAMIC EQUATIONS OF GENERAL RELATIVITY

Evaluate the final indicated source terms in (21.115) from the Lagrangian (21.100) of Maxwell electrodynamics, regarded as a function of the A_i and the g_{ij}.

Exercise 21.14. THE CHOICE OF ϕ DOESN'T MATTER

Prove the statement in the text that the dynamic development of the electric and magnetic fields themselves is independent of the choice made for the scalar potential $\phi(t, x, y, z)$ in the analysis (a) in flat spacetime in Minkowski coordinates and (b) in general relativity, according to equations (21.103), and (21.107) as generalized in exercise 21.12.

Exercise 21.15. THE CHOICE OF SLICING OF SPACETIME DOESN'T MATTER

Given a metric ${}^{(3)}g_{ij}(x, y, z)$ and an extrinsic curvature $K^{ij}(x, y, z)$ on a spacelike hypersurface $\boldsymbol{\Sigma}$, and given that these quantities satisfy the initial-value equations (21.116) and (21.117), and given two alternative choices for the lapse and shift functions (N, N_i) and $(N + \delta N, N_i + \delta N_i)$, show that the curvature itself (as distinguished from its components in these two distinct coordinate systems), as calculated at a point $\mathscr{P}$ a "little way" (first order of small quantities) off the hypersurface, by way of the dynamic equations (21.114) and (21.115), is independent of this choice of lapse and shift.

§21.8. INTEGRATING FORWARD IN TIME

In the Hamiltonian formalism of Arnowitt, Deser, and Misner [see also the many papers by many workers on the quantization of general relativity—primarily putting Einstein's theory into Hamiltonian form—cited, for example, in references 1 and 2 of Wheeler (1968)], the dynamics of geometry takes a form quite similar to the Hamiltonian dynamics of geometry. There one gives x and p at a starting time and integrates two first-order equations for dx/dt and dp/dt ahead in time to find these dynamically conjugate variables at all future times. Here one gives appropriate values of g_{ij} and π^{ij} over an initial spacelike hypersurface and integrates the two first-order equations (21.114) and (21.115) ahead in time to find the geometry at future times. For example, one can rewrite the differential equations as difference equations according to the practice by now familiar in modern hydrodynamics, and then carry out the integration on an electronic digital computer of substantial memory capacity.

Time in general relativity has a many-fingered quality very different from the one-parameter nature of time in nonrelativistic particle mechanics [see, however, Dirac, Fock, and Podolsky (1932) for a many-time formalism for treating the relativistic dynamics of a system of many interacting particles]. He who is studying the geometry is free to push ahead the spacelike hypersurface faster at one place than another, so long as he keeps it spacelike. This freedom expresses itself in the lapse function $N(t, x, y, z)$ at each stage, t, of the integration. Equations (21.114) and (21.115) are not a conduit to feed out information on N to the analyst. They are a conduit for the analyst to feed in information on N. The choice of N is to be made, not by nature, but by man. The dynamic equations cannot begin to fulfill their purpose until this choice is made. The "time parameter" t is only a label to distinguish one spacelike hypersurface from another in a one-parameter family of hypersurface; but N thus tells the spacing in proper time, as it varies from place to place, between the successive slices on which one chooses to record the time-evolution of the geometry. A cinema camera can record what happens only one frame at a time, but the operator can make a great difference in what that camera sees by his choice of angle for the filming of the scene. So here, with the choice of slicing.

Lapse and shift chosen to push forward the integration in time as one finds most convenient

Another choice is of concern to the analyst, especially one doing his analysis on a digital computer. He is in the course of determining, via (21.114–21.115) written as difference equations, what happens on a lattice work of points, typified by $x = \ldots, 73, 74, 75, 76, 77, \ldots$, etc. He finds that the curvatures are developing most strongly in a localized region in the range around $x = 83$ to $x = 89$. He wants to increase the density of coverage of his tracer points in this region. He does so by causing points at lesser and greater x values to drift into this region moment by moment as t increases: $t = \ldots, 122, 123, 124, \ldots$. He makes the tracer points at lesser x-values start to move to the right (N_1 positive) and points at greater x-values move to the left (N_1 negative). In other words, the choice of the three shift functions $N_i(t, x, y, z)$ is just as much the responsibility of the analyst as is the choice of the lapse function N. The equations will never tell him what to pick. He has to tell the equations.

These options, far from complicating dynamic equations (21.114–21.115), make them flexible and responsive to the wishes of the analyst in following the course of whatever geometrodynamic process is in his hands for study.

The freedom that exists in general relativity in the choice of the four functions N, N_i, is illuminated from another side by comparing it with the freedom one has in electrodynamics to pick the one function $\phi(t, x, y, z)$, the scalar potential. In no way do the dynamic Maxwell equations (21.103) and (21.107), as generalized in exercise 21.12 determine ϕ. Instead they demand that it be determined (by the analyst) as the price for predicting the time-development of the vector potential A_i. An altered choice of $\phi(t, x, y, z)$ in its dependence on position and time means altered results from the dynamic equations for the development of the three A_i in time and space. However, the physically significant quantities, the electric and magnetic fields themselves on successive hypersurfaces, come out the same (exercise 21.14) regardless of this choice of ϕ. Similarly in geometrodynamics: an altered choice for the four

Same 4-geometry regardless of lapse and shift options

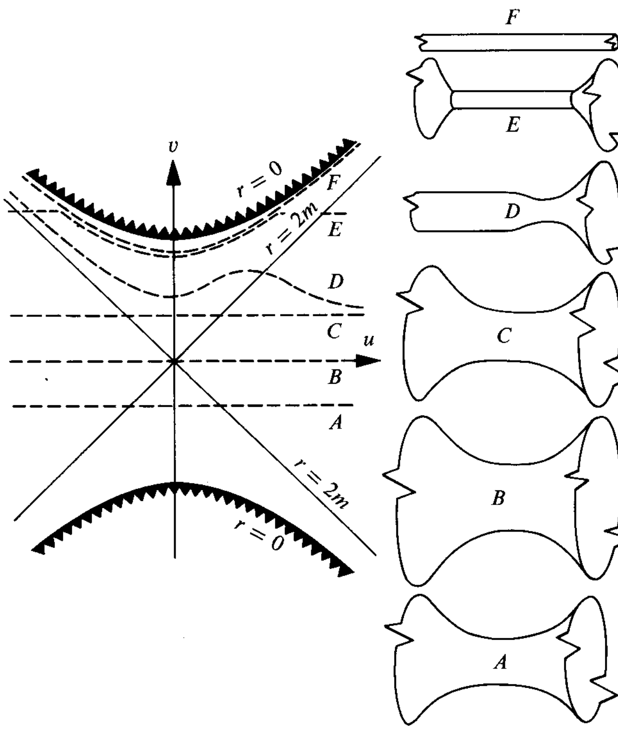


Figure 21.4.
Some of the many ways to make distinct spacelike slices through one and the same $^{(4)}\mathscr{G}$, the complete Schwarzschild 4-geometry.

functions N, N_i, means (a) an altered laying down of coordinates in spacetime, and therefore (b) altered results for the intrinsic metric $^{(3)}g_{ij}$ and extrinsic curvature K^{ij} of successive spacelike hypersurfaces, but yields the same 4-geometry $^{(4)}\mathscr{G}$ (Figure 21.4) regardless of this choice of coordinatization (exercise 21.15).

§21.9. THE INITIAL-VALUE PROBLEM IN THE THIN-SANDWICH FORMULATION

Initial-value data: what is freely disposable? and what is thereby fixed?

Given appropriate initial-value data, one can integrate the dynamic equations ahead in time and determine the evolution of the geometry; but what are "appropriate initial-value data"? They are six functions $^{(3)}g_{ij}(x, y, z)$ plus six more functions $\pi^{ij}(x, y, z)$ or $K^{ij}(x, y, z)$ that together satisfy the four initial-value equations (21.116) and (21.117). To be required to give coordinates and momenta accords with the familiar plan of Hamiltonian mechanics; but to have consistency conditions or "constraints" imposed on such data is less familiar. A particle moving in two-dimensional space is catalogued by coordinates x, y, and coordinates p_x, p_y; but a particle forced to remain on the circle $x^2 + y^2 = a^2$ satisfies the constraint $xp_x + yp_y = 0$. Thus the existence of a "constraint" is a signal that the system possesses fewer degrees of freedom than one would otherwise suppose. Fully to analyze the four "initial-value" or "constraint" conditions (21.116) and (21.117) is thus to determine (1) how many dynamic degrees of freedom the geometry possesses and (2) what these degrees of freedom are; that is to say, precisely what "handles" one can freely adjust to govern completely the geometry and its evolution with time. The counting one can do today, with the conclusion that the geometry possesses the same count of true degrees of freedom as the electromagnetic field. The identification of the "handles," or freely adjustable features of the dynamics, is less advanced for geometry than it is for electromagnetism (Box 21.2), but most instructive so far as it goes.

By rights the identification of the degrees of freedom of the field, whether that of Einstein or that of Faraday and Maxwell, requires nothing more than knowing what must be fixed on initial and final spacelike hypersurfaces to make the appropriate variation principle well-defined. One then has the option whether (1) to give that quantity on both hypersurfaces or (2) to give that quantity and its dynamic conjugate on one hypersurface or (3) to give the quantity on both hypersurfaces, as in (1), but go to the limit of an infinitely thin sandwich, so that one ends up specifying the quantity and its time rate of change on one hypersurface. This third "thin sandwich" procedure is simplest for a quick analysis of the initial-value problem in both electrodynamics and geometrodynamics. Take electrodynamics first, as an illustration.

In electromagnetism, give magnetic field and its rate of change as initial data

Give the divergence-free magnetic field and its time-rate of change: on an arbitrary smooth spacelike hypersurface in curved spacetime in the general case; on the hypersurface $t = 0$ in Minkowski spacetime in the present illustrative treatment,

$$\mathcal{B}^i(0, x, y, z) \text{ given,} \tag{21.118}$$

$$\dot{\mathcal{B}}^i(0, x, y, z) = \left(\frac{\partial \mathcal{B}^i}{\partial t}\right) \text{ also given.} \tag{21.119}$$

These quantities together contain four and only four independent data per space point. How is one now to obtain the momenta $\pi^i \sim -\mathcal{E}^i$ so that one can start integrating the dynamic equations (21.103) and (21.107) forward in time? (1) Find a set of three functions $A_i(0, x, y, z)$ such that their curl gives the three specified $\mathcal{B}^i$. That this can be done at all is guaranteed by the vanishing of the divergence $\mathcal{B}^i{}_{,i}$. However, the choice of the A_i is not unique. The new set of potentials $A_{i\text{new}} = A_i + \partial\lambda/\partial x^i$ with arbitrary smooth λ, provide just as good a solution as the original A_i. No matter. Pick one solution and stick to it. (2) Similarly, find a set of three $\dot{A}_i(0, x, y, z)$ such that their curl gives the specified $\dot{\mathcal{B}}^i(0, x, y, z)$, and resolve all arbitrariness of choice by *fiat*. (3) Recall that the electric field (negative of the field momentum) is given by

$$\mathcal{E}_i = -\dot{A}_i - \partial\phi/\partial x^i \tag{21.120}$$

(formula valid without amendment only in flat space). The initial-value or constraint equation $\mathcal{E}^i{}_{,i} = 0$ translates to the form

$$\nabla^2\phi = -\eta^{ij}\dot{A}_{i,j}. \tag{21.121}$$

Box 21.2 COUNTING THE DEGREES OF FREEDOM OF THE ELECTROMAGNETIC FIELD

A. First Approach: Number of "Field Coordinates" per Spacepoint

Superficial tally of the degrees of freedom of the source-free electromagnetic field gives three field coordinates $A_i(x, y, z)$ per spacepoint on the initial simultaneity Σ, plus three field momenta $\pi^i_{\text{true}} = \pi^i/4\pi$ [with $\pi^i = -\mathcal{E}^i(x, y, z)$] per spacepoint.

Closer inspection reveals that the number of coordinate degrees of freedom per spacepoint is not three but two. Thus the change in vector potential $A_i \longrightarrow A_i + \partial\lambda/\partial x^i$ makes no change in the actual physics, the magnetic field components,

$$B^i = \frac{1}{2}[ijk](\partial A_k/\partial x^j - \partial A_j/\partial x^k).$$

Moreover, though those components are three in number, they satisfy one condition per spacepoint, $\mathcal{B}^i{}_{,i} = 0$, thus reducing the effective net number of coordinate degrees of freedom per spacepoint to two.

The momentum degrees of freedom per spacepoint are likewise reduced from three to two by the one condition per spacepoint $\mathcal{E}^i{}_{,i} = 0$.

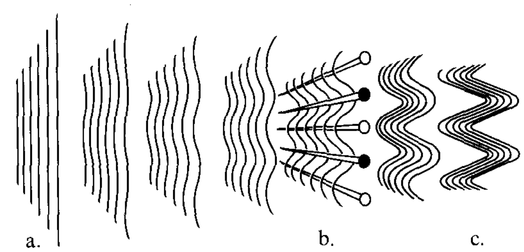
B. Alternative Approach: Count Fourier Coefficients

In textbooks on field theory [see, for example, Wentzel (1949)], attention focuses on flat spacetime. The electromagnetic field is decomposed by Fourier analysis into individual running waves. Instead of counting degrees of freedom per point in coordinate space, one does the equivalent: counts up degrees of freedom per point in wave-number space. Thus for each (k_x, k_y, k_z), there are two independent states of polarization. Each state of polarization requires for its description an amplitude ("coordinate") and time-rate of change of amplitude ("momentum") at the initial time, t_0'. Thus the number of degrees of freedom per point in wave-number space is two for coordinates and two for momenta, in accord with what one gets by carrying out the count in coordinate space.

In curved spacetime, Fourier analysis is a less convenient way of identifying the degrees of freedom of the electromagnetic field [for such a Fourier analysis, see Misner and Wheeler (1957), especially their Table X and following text] than direct analysis in space, as above.

C. Another Alternative: Analyze "Deformation of Structure"

Still a third way to get a handle on the degrees of freedom of a divergence-free field, whether $\mathcal{E}$ or $\mathcal{B}$, rests on the idea of deformation of structure [diagram from Wheeler (1964)]. Represent the



magnetic field by Faraday's picture of lines of force (a) continuing through space without ever ending, automatic guarantee that $\mathcal{B}^i{}_{,i}$ is everywhere zero. Insert "knitting needles" (b) into the spaghetti-like structure of the lines of force and move these needles as one will. Sliding the "knitting needles" along a line of force causes no movement of the line of force. (c) With the help of two knitting needles perpendicular to each other and to the line of force, one can give any given line of force any small displacement one pleases perpendicular to its length: again two degrees of freedom per spacepoint. Granted any non-zero field to begin with, no matter how small, one can build it up by a sequence of such small deformations to agree with any arbitrary field pattern of zero divergence, no matter what its complexity and strength may be.

Solve for ϕ. Then (4) equation (21.120) gives the initial-value electric field, or electrodynamic field momentum $\pi^i \sim -\mathcal{E}^i$, required (along with the field coordinate A_i) for starting the integration of the dynamic equations (21.103) and (21.107). [Misner and Wheeler (1957) deal with the additional features that come in when the space is multiply connected. Each wormhole or handle of the geometry is able to trap electric lines of force. The flux trapped in any one wormhole defines the classical electric charge q_w associated with that wormhole. One has to specify all these charges once and for all in addition to the data (21.118) and (21.119) in order to determine fully the dynamic evolution of the electromagnetic field. There is no geometrodynamic analog to electric charge, according to Unruh (1971).] (5) In this integration, the scalar potential ϕ at each subsequent time step is not to be calculated; it is to be chosen. Only when one has made this free choice definite do the dynamic equations come out with definite results for the A_i and the π^i or $\mathcal{E}^i$ at these successive steps.

Scalar potential: fixed at start; freely disposable later

In the thin-sandwich formulation of the initial-value problem of electrodynamics, to summarize, one gives $\mathcal{B}^i$ and $\dot{\mathcal{B}}^i$ (equivalent to $\boldsymbol{\mathcal{B}}$ on two nearby hypersurfaces). One chooses the A_i and $\dot{A}_i$ with much arbitrariness to represent these initial-value data. The arbitrariness having been seized on to give the initial A_i and $\dot{A}_i$, there is no arbitrariness left in the initial ϕ. However, at all subsequent times the situation is just the other way around. All the arbitrariness is sopped up in the choice of the ϕ, leaving no arbitrariness whatever in the three A_i (as given by the integration of the dynamic equation).

The situation is quite similar in geometrodynamics. One gives the beginnings of a 1-parameter family of spacelike hypersurfaces; namely,

In ADM treatment, give 3-geometry and its time-rate of change

$$^{(3)}\mathcal{G}(0) \text{ given,} \tag{21.122}$$

$$^{(3)}\dot{\mathcal{G}}(0) = \frac{\partial\,^{(3)}\mathcal{G}}{\partial t} \text{ given,} \tag{21.123}$$

Then (1) one picks a definite set of coordinates $x^i = (x, y, z)$ and in terms of those coordinates finds the unique metric coefficients $g_{ij}(x, y, z)$ that describe that 3-geometry. The existence of a solution is guaranteed by the circumstance that $^{(3)}\mathcal{G}$ is a Riemannian geometry. However, one could have started with different coordinates and ended up with different metric coefficients for the description of the same 3-geometry. No matter. Pick one set of coordinates, take the resulting metric coefficients, and stick to them as giving half the required initial-value data. (2) Similarly, to describe the 3-geometry $^{(3)}\mathcal{G} + {}^{(3)}\dot{\mathcal{G}}\, dt$ at the value of the parameter $t + dt$, make use of coordinates $x^i + \dot{x}^i\, dt$ and arrive at the metric coefficients $g_{ij} + \dot{g}_{ij}\, dt$. The arbitrariness in the x^i having thus been resolved by *fiat*, and the $^{(3)}\mathcal{G}$ being given as definite initial physical data, the g_{ij} are thereby completely fixed. (3) Recall that the components of the extrinsic curvature K_{ij} or the momenta π^{ij} are given in terms of the g_{ij} and $\dot{g}_{ij}$ and the lapse and shift functions N and N_i by (21.67) or by (21.67) plus (21.91) or by (21.114). The four initial-value or "constraint" equations (21.116) and (21.117) thus become four conditions for finding the four quantities N, N_i. One can shorten the writing of these conditions by introducing the abbreviations

$$\gamma_{ij} = \frac{1}{2}[N_{i|j} + N_{j|i} - \partial g_{ij}/\partial t] \tag{21.124}$$

and

$$\gamma_2 = \begin{pmatrix}\text{“shift} \\ \text{anomaly”}\end{pmatrix} = (\mathrm{Tr}\,\boldsymbol{\gamma})^2 - \mathrm{Tr}\,\boldsymbol{\gamma}^2 \tag{21.125}$$

(both for functions of x, y, z on the initial simultaneity). Then one has

$$^{(3)}R + \gamma_2/N^2 = 16\pi T_{nn} = 16\pi T^{nn} \tag{21.126}$$

for the one initial-value equation; and for the other three,

$$\left[\frac{\gamma_i{}^k - \delta_i{}^k \mathrm{Tr}\,\gamma}{N}\right]_{|k} = -8\pi T_i{}^n. \tag{21.127}$$

Lapse and shift initially determinate; thereafter freely disposable

In summary, one chooses the g_{ij} and $\dot{g}_{ij}$ with much arbitrariness (because of the arbitrariness in the coordinates, not by reason of any arbitrariness in the physics) to represent the given initial-value data, $^{(3)}\mathscr{G}$ and $^{(3)}\dot{\mathscr{G}}$. The arbitrariness at the initial time all having been soaked up in this way, one expects no arbitrariness to be left in the initial N and N_i as obtained by solving (21.126) and (21.127). However, on all later spacelike slices, the award of the arbitrariness is reversed. The lapse and shift functions are freely disposable, but, with them once chosen, there is no arbitrariness whatever in the six g_{ij} (and the six K^{ij} or π^{ij}) as given by the integration of the dynamic equations (21.114) and (21.115). The analogy with electrodynamics is clear. There the one "gauge-controlled" function ϕ was fixed at the start by the elliptic equation (21.121), but was thereafter free. Here the four lapse and shift functions are fixed at the start by the four equations (21.126) and (21.127), but are thereafter free.

Exercise 21.16 applies the initial-value equation (21.126) to analyze the whole evolution in time of any Friedmann universe in which one knows the equation $p = p(\rho)$ connecting pressure with density. Exercise 21.17 looks for a variation principle on the spacelike hypersurface Σ equivalent in content to the elliptic initial-value equation (21.121) for the scalar potential ϕ. Exercises 21.18 and 21.19 look for similar variation principles to determine the lapse and shift functions.

Counting initial-value data

How many degrees of freedom, or how many "handles," are there in the specification of the 4-geometry that one will obtain? The metric coefficients of the initial 3-geometry provided six numbers per space point. However, they were arbitrary to the extent of a coordinate transformation, specified by three functions of position,

$$\begin{aligned} x &= x(x', y', z'), \\ y &= y(x', y', z'), \\ z &= z(x', y', z'). \end{aligned}$$

The net number of quantities per space point with any physical information was therefore $6 - 3 = 3$. One can visualize these three functions as the three diagonal components of the metric in a coordinate system in which g_{ij} has been transformed to diagonal form. Ordinarily it is not useful to go further and actually spell out the analysis in any such narrowly circumscribed coordinate system.

Now think of the ${}^{(3)}\mathscr{G}$ in question as imbedded in the ${}^{(4)}\mathscr{G}$ that comes out of the integrations. Moreover, think of that ${}^{(4)}\mathscr{G}$ as endowed with the lumps, bumps, wiggles, and waves that distinguish it from other generic 4-geometries and that make Minkowski geometry and special cosmologies so unrepresentative. The ${}^{(3)}\mathscr{G}$ is a slice in that ${}^{(4)}\mathscr{G}$. It partakes of the lumps, bumps, wiggles, and waves present in all those regions of the ${}^{(4)}\mathscr{G}$ that it intersects. To the extent that the ${}^{(4)}\mathscr{G}$ is generic, it does not allow the ${}^{(3)}\mathscr{G}$ to be moved to another location without becoming a different ${}^{(3)}\mathscr{G}$. If one tries to push the ${}^{(3)}\mathscr{G}$ "forward in time" a little in a certain locality, leaving it unchanged in location elsewhere, one necessarily changes the ${}^{(3)}\mathscr{G}$. By this circumstance, one sees that the ${}^{(3)}\mathscr{G}$ "carries information about time" [Sharp (1960); Baierlein, Sharp, and Wheeler (1962)]. Moreover, this "forward motion in time" demands for its description one number per space point. It is possible to think of this number in concrete terms by imagining an arbitrary coordinate system $\bar{t}$, $\bar{x}$, $\bar{y}$, $\bar{z}$ laid down in the ${}^{(4)}\mathscr{G}$. Then the hypersurface can be conceived as defined by the value $\bar{t} = \bar{t}(\bar{x}, \bar{y}, \bar{z})$ at which it cuts the typical line $\bar{x}$, $\bar{y}$, $\bar{z}$. A forward movement carries it to $\bar{t}(\bar{x}, \bar{y}, \bar{z}) + \delta\bar{t}(\bar{x}, \bar{y}, \bar{z})$, and changes shape and metric coefficients on ${}^{(3)}\mathscr{G}$ accordingly. It is usually better not to tie one's thinking down to such a concrete model, but rather to recognize as a general point of principle (1) that the location of the ${}^{(3)}\mathscr{G}$ in spacetime demands for its specification one datum per spacepoint, and (2) that this datum is already willy-nilly present in the three data per spacepoint that mark any ${}^{(3)}\mathscr{G}$.

Four pieces of geometrodynamic information per space point on initial simultaneity

In conclusion, there are only two data per spacepoint in a ${}^{(3)}\mathscr{G}$ that really tell anything about the ${}^{(4)}\mathscr{G}$ in which it is imbedded, or to be imbedded (as distinguished from where the ${}^{(3)}\mathscr{G}$ slices through that ${}^{(4)}\mathscr{G}$). Similarly for the other ${}^{(3)}\mathscr{G}$ that defines the other "face of the sandwich," whether thick or thin. Thus one concludes that the specification of ${}^{(3)}\mathscr{G}$ and ${}^{(3)}\dot{\mathscr{G}}$ actually gives four net pieces of dynamic information per spacepoint about the ${}^{(4)}\mathscr{G}$ (all the rest of the information being "many-fingered time," telling where the 3-geometries are located in that ${}^{(4)}\mathscr{G}$). According to this line of reasoning, geometrodynamics has the same number of dynamic degrees of freedom as electrodynamics. One arrives at the same conclusion in quite another way through the weak-field analysis (§35.3) of gravitational waves on a flat spacetime background: the same ranges of possible wave numbers as for Maxwell waves; and for each wave number two states of polarization; and for each polarization one amplitude and one phase (the equivalent of one coordinate and one momentum).

In electrodynamics in a prescribed spacetime manifold, one has a clean separation between the one time-datum per spacepoint (when one deals with electromagnetism in the context of many-fingered time) and the two dynamic variables per spacepoint; but not so in the superspace formulation of geometrodynamics. There the two kinds of quantities are inextricably mixed together in the one concept of 3-geometry.

Turn from initial- and final-value data to the action integral that is determined by (1) these data and (2) the principle that the action be an extremum,

$$I = I_{\text{extremum}} = S.$$

The action depends on the variables on the final hypersurface, according to the formula

$$S = S(\Sigma, \boldsymbol{B}) \tag{21.128}$$

in electrodynamics, but according to the formula

$$S = S({}^{(3)}\mathcal{G}) \tag{21.129}$$

in geometrodynamics. In each case, there are three numbers per spacepoint in the argument of the functional (one in Σ; two in a divergence-free magnetic field; three in ${}^{(3)}\mathcal{G}$).

Problem in assuring completeness and consistency of initial data

This mixing of the one many-fingered time and the two dynamic variables in a 3-geometry makes it harder in general relativity than in Maxwell theory to know when one has in hand appropriate initial value data. Give Σ and give $\mathcal{B}$ and $\dot{\mathcal{B}}$ on Σ: that was enough for electrodynamics. For geometrodynamics, to give the six $g_{ij}(x, y, z)$ and the six $\dot{g}_{ij}(x, y, z)$ is not necessarily enough. For example, let the time parameter t be a fake, so that dt, instead of leading forward from a given hypersurface Σ to a new hypersurface $\Sigma + d\Sigma$, merely recoordinatizes the present hypersurface:

$$\begin{aligned} x^i &\longrightarrow x^i - \xi^i\, dt, \\ g_{ij} &\longrightarrow g_{ij} + (\xi_{i|j} + \xi_{j|i})\, dt. \end{aligned} \tag{21.130}$$

A first inspection may make one think that one has adequate data in the six g_{ij} and the six

$$\dot{g}_{ij} = \xi_{i|j} + \xi_{j|i}, \tag{21.131}$$

but in the end one sees that one has not both faces of the thin sandwich, as required, but only one. Thus one must reject, as improperly posed data in the generic problem of dynamics, any set of six $\dot{g}_{ij}$ that let themselves be expressed in the form (21.131) [Belasco and Ohanian (1969)].

Similar difficulties occur when the two faces of the thin sandwich, instead of coinciding everywhere, coincide in a limited region, be it three-dimensional, two-dimensional, or even one-dimensional ("crossover of one face from being earlier than the other to being later"). Thus it is enough to have (21.131) obtaining even on only a curved line in Σ to reject the six g_{ij} as inappropriate initial-value data.

The "thin sandwich conjecture"

That one can impose conditions on the g_{ij} and $\dot{g}_{ij}$ which will guarantee existence and uniqueness of the solution $N(x, y, z)$, $N_i(x, y, z)$ of the initial-value equations (21.126) and (21.127) is known as the "thin-sandwich conjecture," a topic on which there has been much work by many investigators, but so far no decisive theorem.

To presuppose existence and uniqueness is to make the first step in giving mathematical content to Mach's principle that the distribution of mass-energy throughout space determines inertia (§21.12).

§21.10. THE TIME-SYMMETRIC AND TIME-ANTISYMMETRIC INITIAL-VALUE PROBLEMS

Turn from the general initial-value problem to two special initial-value problems that lend themselves to detailed treatment, one known as the time-symmetric initial-value problem, the other as the time-antisymmetric problem.

A 4-geometry is said to be time-symmetric when there exists a spacelike hypersurface Σ at all points of which the extrinsic curvature vanishes. In this case the three initial value equations (21.127) are automatically satisfied, and the fourth reduces to a simple requirement on the three-dimensional scalar curvature invariant,

$$R = 16\pi\rho. \tag{21.132}$$

Still further simplifications result when one limits attention to empty space. Simplest of all is the case of spherical symmetry in which (21.132) yields at once the full Schwarzschild geometry at the moment of time symmetry (two asymptotically flat spaces connected by a throat), as developed in exercise 21.20.

Consider a 3-geometry with metric

$$ds_1{}^2 = g_{(1)ik}\, dx^i\, dx^k. \tag{21.133}$$

Call it a "base metric." Consider another 3-geometry with metric

$$ds_2{}^2 = \psi^4(x^i)\, ds_1{}^2. \tag{21.134}$$

Angles are identical in the two geometries. On this account they are said to be conformally equivalent. The scalar curvature invariants of the two 3-geometries are related by the formula [Eisenhart (1926)]

$$R_2 = -8\psi^{-5}\, \nabla_1{}^2\psi + \psi^{-4}R_1, \tag{21.135}$$

where

$$\nabla_1{}^2\psi = \psi_{|i}{}^{|i} = g_1{}^{-1/2}(\partial/\partial x^i)[g_1{}^{1/2}g^{ik}(\partial\psi/\partial x^k)] \tag{21.136}$$

Demand that the scalar curvature invariant R_2 vanish, and arrive [Brill (1959)] at the "wave equation"

$$\nabla_1{}^2\psi - (R_1/8)\psi = 0 \tag{21.137}$$

for the conformal correction factor ψ. Brill takes the base metric to have the form suggested by Bondi,

$$ds_1{}^2 = e^{2Aq_1(\rho,z)}(dz^2 + d\rho^2) + \rho^2\, d\phi^2, \tag{21.138}$$

and takes the conformal correction factor ψ also to possess axial symmetry. In the application:

$q_1(\rho, z)$ measures the "distribution of gravitational wave amplitude," assumed for simplicity to vanish outside $r = (\rho^2 + z^2)^{1/2} = a$;

A measures the "amplitude of the distribution of gravitational wave amplitude";

$\psi(\rho, z)$ is the conformal correction factor, which varies with position at large distances as $1 + (m/2r)$. The quantity m(cm) is uniquely determined by the condition that the geometry be asymtotically flat. It measures the mass-energy of the distribution of gravitational radiation.

Wave amplitude determines mass-energy: $m = m(A)$

The mass m of the gravitational radiation is proportional to A^2 for small values of the amplitude A. It is inversely proportional to the reduced wavelength $\lambda\!\!\!^{-} =$ (effective wavelength/2π) that measures the scale of rapid variations in the gravitational wave amplitude $q_1(\rho, z)$ in the "active zone." Thus the metric is dominated by wiggles, proportional in amplitude to A, in the active zone, and at larger distances dominated by something close to a Schwarzschild $(1 + 2m/r)$ factor in the metric. When the amplitude A is increased, a critical value is attained, $A = A_{\text{crit}}$, at which m goes to infinity and the geometry curves up into closure ("universe closed by its own content of gravitational-wave energy"). Further analysis and examples will be found in Wheeler (1964a), pp. 399–451, also in Wheeler (1964c).

"Time-antisymmetric" initial-value data

Brill has carried out a similar analysis [Brill (1961)] for the vacuum case of what he calls time-antisymmetric initial-value conditions, sketched below as amended by York (1973). (1) The initial slice is maximal, Tr $\boldsymbol{K} = 0$. (2) This slice is conformally flat,

$$g_{ij} = \psi^4\, \delta_{ij}. \tag{21.139}$$

(3) Work in the "base space" with metric δ_{ij} and afterwards transform to the geometry (21.139). Three of the initial-value equations become

$$K^{ij}_{\text{base}, j} = 0. \tag{21.140}$$

To solve these equations, (1) take any localized trace-free symmetric tensor B_{km}; (2) solve the flat-space Laplace equation $\nabla^2 A = (3/2)\, \partial^2 B_{km}/\partial x^k\, \partial x^m$ for A; (3) define the six potentials $A_{km} = B_{km} + \frac{1}{3}A\, \delta_{km}$; and (4) calculate

$$K^{ij}_{\text{base}} = [ik\ell][jmn]\, \partial^2 A_{km}/\partial x^\ell\, \partial x^n, \tag{21.141}$$

that automatically satisfy (21.140) and give Tr $\boldsymbol{K}_{\text{base}} = 0$. Then $K^{ij} = \psi^{-10} K^{ij}_{\text{base}}$ *also* automatically satisfies these conditions, but now in the *curved* geometry (21.139). The final initial-value equation becomes a quasilinear elliptic equation, in the flat base space, for the conformal factor ψ,

$$8\nabla^2_{\text{base}}\, \psi + \psi^{-7} \sum_{i,j} (K_{\text{base }ij})^2 = 0. \tag{21.142}$$

The asymptotic form of ψ reveals that the mass of the wave is positive.

In addition to the time-symmetric and time-antisymmetric cases, there are at least two further cases where the initial-value problem possess special simplicity. One is the case of a geometry endowed with a symmetry, as, for example, for the Friedmann universe of Chapter 27 or the mixmaster universe of Chapter 30 or cylindrical gravitational waves in the treatment of Kuchař (1971a). One starts with a spacelike slice on which the g_{ij} and π^{ij} have a special symmetry, and makes all future spacelike slices in a way that preserves this symmetry. The geometry on any one of these simultaneities, though almost entirely governed by these symmetry considerations, still typically demands some countable number of parameters for its complete determination, such as the radius of the Friedmann universe, or the three principal radii of curvature of the mixmaster universe. These parameters and the momenta conjugate to them define a miniphase space. In this miniphase space, the dynamics runs its course as for any other problem of classical dynamics [see, for example, Box 30.1 and Misner (1969) for the mixmaster universe; Kuchař (1971a) and (1972) for waves endowed with cylindrical symmetry; and Gowdy (1973) for waves with spherical symmetry]. Even the evidence for the existence of many-fingered time, most characteristic feature of general relativity, is suppressed as the price for never having to give attention to any spacelike slice that departs from the prescribed symmetry.

Finite dimensional dynamics for geometries endowed with high symmetry

---

EXERCISES

Exercise 21.16. POOR MAN'S WAY TO DO COSMOLOGY

Consider a spacetime with the metric

$$ds^2 = -dt^2 + a^2(t)[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)],$$

corresponding to a 3-geometry with the form of a sphere of radius $a(t)$ changing with time. Show that the tensor of extrinsic curvature as expressed in a local Euclidean frame of reference is

$$\boldsymbol{K} = -a^{-1}(da/dt)\boldsymbol{1},$$

where $\boldsymbol{1}$ is the unit tensor. Show that the initial value equation (21.77) reduces to

$$(6/a^2)(da/dt)^2 + (6/a^2) = 16\pi\rho(a)$$

[for the value of the second term on the left, see exercise 14.3 and Boxes 14.2 and 14.5], and explain why it is appropriate to write the term on the right as $6a_0/a^3$ for a "dust-filled model universe." More generally, given any equation of state, $p = p(\rho)$, explain how one can find $\rho = \rho(a)$ from

$$d(\rho a^3) = -p\, d(a^3);$$

and how one can thus forecast the history of expansion and recontraction, $a = a(t)$.

Exercise 21.17. THIN-SANDWICH VARIATIONAL PRINCIPLE FOR THE SCALAR POTENTIAL IN ELECTRODYNAMICS

(a) Choose the unknown U^m in the expression

$$\frac{1}{8\pi} g^{mn} \frac{\partial\phi}{\partial x^m}\frac{\partial\phi}{\partial x^n} + U^m \frac{\partial\phi}{\partial x^n}$$

in such a way that this expression, multiplied by the volume element $g^{1/2}\,d^3x$, and integrated over the simultaneity Σ, is extremized by a ϕ, and only by a ϕ, that satisfies the initial-value equation (21.108) of electrodynamics.

(b) Show that the resulting variational principle, instead of having to be invented "out of the blue," is none other than what follows directly from the action principle build on the Lagrangian density (21.100) of electrodynamics (independent variation of ϕ and the three A_i everywhere between the two faces of a sandwich to extremize I, subject only to the prior specification of the A_i on the two faces of the sandwich, in the limit where the thickness of the sandwich goes to zero).

Exercise 21.18. THIN-SANDWICH VARIATIONAL PRINCIPLE FOR THE LAPSE AND SHIFT FUNCTIONS IN GEOMETRODYNAMICS

(a) Extremize the action integral

$$\begin{aligned} I_3 = \int \{ & [R - (\mathrm{Tr}\boldsymbol{K})^2 + \mathrm{Tr}\boldsymbol{K}^2 - 2T^*_{nn}]N \\ & - 2T^{*k}_{n} N_k \} g^{1/2}\, d^3x \end{aligned}$$

with respect to the lapse and shift functions, and show that one arrives in this way at the four initial-value equations of geometrodynamics. It is understood that one has given the six g_{ij} and the six $\partial g_{ij}/\partial t$ on the simultaneity where the analysis is being done. The extrinsic curvature is considered to be expressed as in (21.67) in terms of these quantities and the lapse and shift. The energy density and energy flow are referred to a unit normal vector $\boldsymbol{n}$ and three arbitrary coordinate basis vectors $\boldsymbol{e}_i$ within the simultaneity, as earlier in this chapter, and the asterisk is an abbreviation for an omitted factor of 8π.

(b) Derive this variational principle from the ADM variational principle by going to the limit of an infinitesimally thin sandwich [see derivation in Wheeler (1964)].

Exercise 21.19. CONDENSED THIN-SANDWICH VARIATIONAL PRINCIPLE

(a) Extremize the action I_3 of the preceding exercise with respect to the lapse function N.

(b) What is the relation between the result and the principle that "3-geometry is a carrier of information about time"?

(c) By elimination of N, arrive at a "condensed thin-sandwich variational principle" in which the only quantities to be varied are the three shift functions N_i.

Exercise 21.20. POOR MAN'S WAY TO SCHWARZSCHILD GEOMETRY

On curved empty space evolving deterministically in time, impose the conditions (1) that it possess a moment of time-symmetry, a spacelike hypersurface, the extrinsic curvature of which, with respect to the enveloping spacetime, is everywhere zero, and (2) that this spacelike hypersurface be endowed with spherical symmetry. Write the metric of the 3-geometry in the form

$$ds^2 = \psi^4(\bar{r})(d\bar{r}^2 + \bar{r}^2\, d\theta^2 + \bar{r}^2 \sin^2\theta\, d\phi^2).$$

From the initial-value equation (21.127), show that the conformal factor ψ up to a multiplicative factor must have the form $\psi = (1 + m/2\bar{r})$. Show that the proper circumference $2\pi\bar{r}\psi^2(\bar{r})$ assumes a minimum value at a certain value of $\bar{r}$, thus defining the *throat* of the 3-geometry. Show that the 3-geometry is mirror-symmetric with respect to reflection in this throat in the sense that the metric is unchanged in form under the substitution $r' = m^2/4\bar{r}$. Find the transformation from the conformal coordinate $\bar{r}$ to the Schwarzschild coordinate r.

§21.11. YORK'S "HANDLES" TO SPECIFY A 4-GEOMETRY

On a simultaneity—or on *the* simultaneity—of extremal proper volume, give the conformal part of the 3-geometry and give the two inequivalent components of the dynamically conjugate momentum in order (1) to have freely specifiable, but also complete, initial-value data and thus (2) to determine completely the whole generic four-dimensional spacetime manifold. This in brief is York's extension (1971, 1972b) to the generic case of what Brill did for special cases (see the preceding section). York and Brill acknowledge earlier considerations of Lichnerowicz (1944) and Bruhat (1962 and earlier papers cited there on conformal geometry and the initial-value problem). But why conformal geometry, and why pick such a special spacelike hypersurface on which to give the four dynamic data per spacepoint?

The degrees of freedom of the geometry in brief

Few solutions of Maxwell's equations are simpler than an infinite plane monochromatic wave in Minkowski's flat spacetime, and few look more complex when examined on a spacelike slice cut through that spacetime in an arbitrary way, with local wiggles and waves, larger-scale lumps and bumps, and still larger-scale general curvatures. No one who wants to explore electrodynamics in its evolution with many-fingered time can avoid these complexities; and no one will accept these complexities of many-fingered time who wants to see the degrees of freedom of the electromagnetic field in and by themselves exhibited in their neatest form. He will pick the simplest kind of timelike slice he can find. On that simultaneity, there are two and only two field coordinates, and two and only two field momenta per spacepoint. Similarly in geometrodynamics.

When one wants to untangle the degrees of freedom of the geometry, as distinct from analyzing the dynamics of the geometry, one therefore retreats from the three items of information per spacepoint that are contained in a 3-geometry [or in any other way of analyzing the geometrodynamics, as especially seen in the "extrinsic time" formulation of Kuchař (1971b and 1972)] and following York (1) picks the simultaneity to have maximal proper volume and (2) on this simultaneity specifies the two "coordinate degrees of freedom per spacepoint" that are contained in the conformal part of the 3-geometry.

Pick hypersurface of extremal proper volume

An element of proper volume $g^{1/2}\,d^3x$ on the spacelike hypersurface Σ undergoes, in the next unit interval of proper time as measured normal to the hypersurface, a fractional increase of proper volume [see Figure 21.3 and equations 21.59 and 21.66] given by

$$-\operatorname{Tr}\boldsymbol{K} = -\frac{1}{2}g^{-1/2}\operatorname{Tr}\boldsymbol{\pi}. \qquad (21.143)$$

For the volume to be extremal this quantity must vanish at every point of Σ. This condition is satisfied in a Friedmann universe (Chapter 27) and in a Taub universe (Chapter 30) at that value of the natural time-coordinate t at which the universe switches over from expansion to recontraction. It is remarkable that the same condition on the choice of simultaneity, Σ, lets itself be formulated in the same natural way,

$$\operatorname{Tr}\boldsymbol{K} = 0 \text{ or } \operatorname{Tr}\boldsymbol{\pi} = 0, \qquad (21.144)$$

for a closed universe altogether deprived of any symmetry whatsoever. Alternatively, one can deal with a spacetime that is topologically the product of an open 3-space by the real line (time). Then it is natural to think of specifying the location in it of a bounding spacelike 2-geometry S with the topology of a 2-sphere. Then one has many ways to fill in the interior of S with a spacelike 3-geometry Σ; but of all these Σ's, only the one that is extremal, or only the ones that are extremal, satisfy (21.144).

Case of open 3-geometry

Who is going to specify this 2-geometry with the topology of a 2-sphere? The choice of that 2-geometry is not a matter of indifference. In a given 4-geometry, distinct choices for the bounding 2-geometry will ordinarily give distinct results for the extremizing 3-geometry, and therefore different choices for the "initial-value simultaneity," Σ. No consideration immediately thrusts itself forward that would give preference to one choice of 2-geometry over another. However, no such infinity of options presents itself when one limits attention to a closed 3-geometry. Therefore it will give concreteness to the following analysis to consider it applied to a closed universe, even though the analysis surely lets itself be made well-defined in an open region by appropriate specification of boundary values on the closed 2-geometry that bounds that open region. In brief, by limiting attention to a closed 3-geometry, one lets the obvious condition of closure take the place of boundary conditions that are not obvious.

York's analysis remains simple when his extrinsic time

$$\tau = \frac{2}{3} g^{-1/2} \operatorname{Tr} \boldsymbol{\pi} = \frac{4}{3} \operatorname{Tr} \boldsymbol{K}$$

has any constant value on the hypersurface, not only the value $\tau = 0$ appropriate for the hypersurface of extremal proper volume.

Meaning of conformal 3-geometry

On the simultaneity Σ specified by the condition of constant extrinsic time, $\tau =$ constant, begin by giving the conformal 3-geometry,

$$< \; = {}^{(3)}< \; = \left(\begin{array}{l}\text{the equivalence class of all those positive definite}\\ \text{Riemannian three-dimensional metrics that are}\\ \text{equivalent to each other under (1) diffeomorphism}\\ \text{(smooth sliding of the points over the mainfold to}\\ \text{new locations) or (2) changes of scale that vary}\\ \text{smoothly from point to point, leaving fixed all}\\ \text{local angles (ratios of local distances), but}\\ \text{changing local distances themselves or (3) both.}\end{array}\right) \tag{21.145}$$

The conformal 3-geometry is a geometric object that lends itself to definition and interpretation quite apart from the specific choice of coordinate system and even without need to use any coordinates at all. The conformal 3-geometry (*on the hypersurface* Σ *where* $\tau =$ *constant*) may be regarded much as one regards the magnetic field in electromagnetism. The case of conformally flat 3-geometry,

$$ds^2 = \psi^4(x, y, z)\, ds^2_{\text{base}} \tag{21.146}$$

(with $g_{ij\,\mathrm{base}} = \delta_{ij}$), is analogous to those initial-value situations in electromagnetism where the magnetic field is everywhere zero (the time-antisymmetric initial-value problem of Brill); but now we consider the case of general ds^2_{base}.

The six metric coefficients g_{ij} of the conformal 3-geometry, subject to being changed by change of the three coordinates x^i, and undetermined at any one point up to a common position-dependent multiplicative factor, carry $6 - 3 - 1 = 2$ pieces of information per spacepoint. In this respect, they are like the components of the divergenceless magnetic field $\mathcal{B}$. The corresponding field momentum $\pi^i_{EM} \propto \mathcal{E}^i$ (Box 21.1, page 496) has its divergence specified by the charge density, and so also carries

$$\text{two pieces of information (in addition to the prescribed information about the density of charge) per spacepoint.} \tag{21.147}$$

The comparison is a little faulty between the components of $\mathcal{B}$ and the metric coefficients. They are more like potentials than like components of the physically relevant field.

The appropriate measure of the "field" in geometrodynamics is the curvature tensor; but how can one possibly define a curvature tensor for a geometry that is as rudimentary as a conformal 3-geometry? York (1971) has raised and answered this question. The Weyl conformal-curvature tensor [equation (13.50) and exercise 13.13] is independent [in the proper $\binom{2}{2}$ representation], in spaces of higher dimensionality, of the position-dependent factor ψ^4 with which one multiplies the metric coefficients, but vanishes identically in three-dimensional space (exercise 21.21). One arrives at a non-zero conformally invariant measure of the curvature only when one goes to one higher derivative (exercise 21.22). In this way, one comes to *York's curvature* $\tilde{\beta}^{ab}$, here called Y^{ab}, a tensor density with these properties:

York's curvature tensor

$$Y^{ab} = Y^{ba} \text{ (symmetric);}$$

$$Y^a_{\ a} = 0 \text{ (traceless);}$$

$$Y^{ab}{}_{|b} = 0 \text{ (transverse);}$$

$$Y^{ab} \text{ invariant with respect to position-dependent changes in the conformal scale factor;}$$

$$Y^{ab} = 0 \text{ when and only when the 3-geometry is conformally flat.} \tag{21.148}$$

Y^{ab} provides what York calls the pure spin-two representation of the 3-geometry intrinsic to Σ. It is the analog of the field $\mathcal{B}$ of electrodynamics on the spacelike initial-value simultaneity. It directly carries physical information about the conformal 3-geometry.

In addition to the conformal geometry ${}^{(3)}\!\!<$, specified by the "potentials" $g_{ij}/g^{1/3}$, and measured by the "field components" Y^{ij}, one must also specify on Σ the corresponding conjugate momenta:

The associated momenta

$$\tilde{\pi}^{ab} = \tilde{\pi}^{ab} \text{ (symmetric)}; \qquad \tilde{\pi}^{a}_{a} = 0 \text{ (traceless)};$$

$\tilde{\pi}^{ab}{}_{|b} = 0$ (transverse) in case there is no flow of energy in space; otherwise

$$\tilde{\pi}^{ab}{}_{|b} = 8\pi \text{ (density of flow of energy)}^{a};$$

two pieces of information (in addition to the prescribed information about the flow of energy) per spacepoint. (21.149)

It might appear to be essential to specify with respect to which of the 3-geometries, distinguished from one another by different values of the conformal factor one calculates the covariant derivatives of tensor densities of weight 5/3 (see §21.2) in (21.148) and (21.149). However, York has shown that the conditions (21.149) do not in any way depend on the value of the conformal factor ψ^4.

These equations (21.149) for what York calls the "momentum density of weight 5/3,"

$$\tilde{\pi}^{ab} = g^{1/3}\left(\pi^{ab} - \frac{1}{3} g^{ab} \operatorname{Tr} \boldsymbol{\pi}\right), \tag{21.150}$$

are linear, and therefore lend themselves to analysis by standard methods. It is a great help in this enterprise that York (1973a,b) has provided a "conformally invariant orthogonal decomposition of symmetric tensors on Riemannian manifolds" that allows one to generate solutions of these requirements ("transverse traceless," "conformal Killing," and "trace" parts, respectively, measure deformation of conformal part of geometry, mere recoordinatization, and change of scale). It is a further assistance, as York notes, that one has the same $\tilde{\pi}^{ab}$ for an entire conformal equivalence class of metrics; that is, for a given

$$\tilde{g}_{ab} = g^{-1/3} g_{ab}, \tag{21.151}$$

no matter how different the g_{ab} and ψ themselves may be.

Unique solution for conformal factor

The conformal 3-geometry and the "momentum density of weight 5/3" once picked, the remaining initial-value equation (21.116) then becomes the "scale" equation,

$$8\nabla^2\psi - {}^{(3)}R\psi + M\psi^{-7} + Q\psi^{-3} - \frac{3}{8}\tau^2\psi^5 = 0, \tag{21.152}$$

for the determination of the conformal factor ψ. Here ∇^2 stands for the Laplacian

$$\nabla^2\psi \equiv g^{-1/2}(\partial/\partial x^a) g^{1/2} g^{ab}(\partial\psi/\partial x^b). \tag{21.153}$$

It, like ${}^{(3)}R$, M, and Q, refers to the base space. It is interesting that

$$\nabla^2 - \frac{1}{8}{}^{(3)}R$$

is a conformally invariant wave operator, whereas ∇^2 itself is not. The quantity M in York's analysis is an abbreviation for

$$M \equiv g^{-5/3} g_{ac} g_{bd} \tilde{\pi}^{ab} \tilde{\pi}^{cd}, \tag{21.154a}$$

and

$$Q \equiv 16\pi\rho_{\text{base}} (= 16\pi\psi^8\rho = 16\pi\psi^8\rho_{\text{in final 3-geometry}}). \tag{21.154b}$$

One seeks a solution ψ that is continuous over the closed manifold and everywhere real and positive. When does such a solution ψ of the elliptic equation (21.152) exist? When is it unique? *Always* (when $M > 0$ and $\tau \neq 0$), is the result of O'Murchadha and York (1973); see also earlier investigations of Choquet-Bruhat (1972). Some of the physical considerations that come into this kind of problem have been discussed by Wheeler (1964a, pp. 370–381).

§21.12. MACH'S PRINCIPLE AND THE ORIGIN OF INERTIA

In my opinion the general theory of relativity can only solve this problem [*of inertia*] *satisfactorily if it regards the world as spatially self-enclosed.*

ALBERT EINSTEIN (1934), p. 52.

On June 25, 1913, two years before he had discovered the geometrodynamic law that bears his name, Einstein (1913b) wrote to Ernst Mach (Figure 21.5) to express his appreciation for the inspiration that he had derived for his endeavors from Mach's ideas. In his great book, *The Science of Mechanics,* Mach [(1912), Chapter 2, section 6] had reasoned that it could not make sense to speak of the acceleration of a mass relative to absolute space. Anyone trying to clear physics of mystical ideas would do better, he reasoned, to speak of acceleration relative to the distant stars. But how can a star at a distance of 10^9 light-years contribute to inertia in the here and the now? To make a long story short, one can say at once that Einstein's theory (1) identifies gravitation as the mechanism by which matter there influences inertia here; (2) says that this coupling takes place on a spacelike hypersurface [in what one, without a closer examination, might mistakenly think to be a violation of the principle of causality; see Fermi (1932) for a discussion and clarification of the similar apparent paradox in electrodynamics; see also Einstein (1934), p. 84: "Moreover I believed that I could show on general considerations a law of gravitation invariant in relation to any transformation of coordinates whatever was inconsistent with the principle of causation. These were errors of thought which cost me two years of excessively hard work, until I finally recognized them as such at the end of 1915"]; (3) supplies in the initial-value equations of geometrodynamics a mathematical tool to describe this coupling; (4) demands closure of the geometry in space [one conjectures; see Wheeler (1959, 1964c) and Hönl (1962)], as a boundary condition on the initial-value equations if they are to yield a well-determined [and, we know now, a unique] 4-geometry; and (5) identifies the collection of local Lorentz frames near any point in this resulting spacetime as what one means quantitatively by speaking of inertia at that point. This is how one ends up with inertia here determined by density and flow of mass-energy there.

No violation of causality, despite appearances

There are many scores of papers in the literature on Mach's principle, including many—even one by Lenin (English translation, 1927)—one could call anti-Machian; and many of them make interesting points [see especially the delightful dialog by Weyl (1924a) on "inertia and the cosmos," and the article (1957) and book (1961) of Sciama]. However, most of them were written before one had anything like the understanding of the initial-value problem that one possesses today. Therefore no

An enormous literature

(*continued on page 546*)

Figure 21.5.
Einstein's appreciation of Mach, written to Ernst Mach June 25, 1913, while Einstein was working hard at arriving at the final November 1915 formulation of standard general relativity. Regarding confirmation at a forthcoming eclipse: "If so, then your happy investigations on the foundations of mechanics, Planck's unjustified criticism notwithstanding, will receive brilliant confirmation. For it necessarily turns out that inertia originates in a kind of interaction between bodies, quite in the sense of your considerations on Newton's pail experiment. The first consequence is on p. 6 of my paper. The following additional points emerge: (1) If one accelerates a heavy shell of matter *S*, then a mass enclosed by that shell experiences an accelerative force. (2) If one rotates the shell relative to the fixed stars about an axis going through its center, a Coriolis force arises in the interior of the shell; that is, the plane of a Foucault pendulum is dragged around (with a practically unmeasurably small angular velocity)." Following the death of Mach, Einstein (1916a) wrote a tribute to the man and his work. Reprinted with the kind permission of the estate of Albert Einstein, Helen Dukas and Otto Nathan, executors.

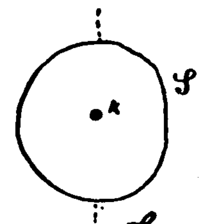
Zürich. 25. VI. 13.

Hoch geehrter Herr Kollege!

Dieser Tage haben Sie wohl meine neue Arbeit über Relativität und Gravitation erhalten, die nach unendlicher Mühe und quälendem Zweifel nun endlich fertig geworden ist. Nächstes Jahr bei der Sonnenfinsternis soll sich zeigen, ob die Lichtstrahlen an der Sonne gekrümmt werden, ob m. a. W. die zugrunde gelegte fundamentale Annahme von der Aequivalenz von Beschleunigung des Bezugssystems einerseits und Schwerefeld andererseits wirklich zutrifft.

Wenn ja, so erfahren Ihre genialen Untersuchungen über die Grundlagen der Mechanik – Planck's ungerechtfertigter Kritik zum Trotz – eine glänzende Bestätigung. Denn es ergibt sich mit Notwendigkeit, dass die <u>Trägheit</u> in einer Art <u>Wechselwirkung</u> der Körper ihren <u>Ursprung</u> habe, ganz im Sinne Ihrer Überlegungen zum Newton'schen Eimer-Versuch.

Eine erste Konsequenz in diesem Sinne finden Sie oben auf Seite 6 der Arbeit. Es hat sich ferner folgendes ergeben:



1) Beschleunigt man eine träge Kugelschale S, so erfährt nach der Theorie ein von ihr eingeschlossener Körper eine beschleunigende Kraft.

2) Rotiert die Schale S um eine durch ihren Mittelpunkt gehende Achse (relativ zum System der Fixsterne („Ruhesystem")), so entsteht im Innern der Schale ein Coriolis-Feld, d. h. ~~das~~ die Ebene des Foucault-Pendels wird (mit einer allerdings praktisch unmessbar kleinen Geschwindigkeit) mitgenommen.

Es ist mir eine grosse Freude, Ihnen dies mitteilen zu können, zumal jene Kritik Plancks mir schon immer höchst ungerechtfertigt erschienen war.

Mit grösster Hochachtung grüsst Sie herzlich

Ihr ergebener A. Einstein.

Ich danke Ihnen herzlich für die Übersendung Ihres Buches.

2

attempt will be made to summarize or analyze the literature, which would demand a book in itself. Moreover, Mach's principle as presented here is more sharply formulated than Einstein ever put it in the literature [except for his considerations arguing that the universe must be closed; see Einstein's book (1950), pp. 107–108]; and Mach would surely have disowned it, for he could never bring himself to accept general (or even special) relativity. Nevertheless, it is a fact that Mach's principle—that matter there governs inertia here—and Riemann's idea—that the geometry of space responds to physics and participates in physics—were the two great currents of thought which Einstein, by means of his powerful equivalence principle, brought together into the present-day geometric description of gravitation and motion.

Mach's principle updated and spelled out

"Specify everywhere the distribution and flow of mass-energy and thereby determine the inertial properties of every test particle everywhere and at all times". Spelled out, this prescription demands (1) a way of speaking about "everywhere": a spacelike hypersurface Σ. Let one insist—in conformity with Einstein—(2) that it be a closed 3-geometry, and for convenience, not out of necessity, (3) that τ be independent of position on Σ. (4) Specify this 3-geometry to the extent of giving the conformal metric; without the specification of at least this much 3-geometry, there would be no evident way to say "where" the mass-energy is to be located. (5) Give density ρ_{base} as a function of position in this conformal 3-geometry. (6) Recognize that giving the mass-energy only of fields other than gravity is an inadequate way to specify the distribution of mass-energy throughout space. Formalistically, to be sure, the gravitational fields does not and cannot make any contribution to the source term that stands on the righthand side of Einstein's field equation. However, the analysis of gravitational waves (Chapters 18 and 35) shows that perturbations in the geometry of scale small compared to the scale of observation have to be regarded as carrying an effective content of mass-energy. Moreover, one has in a geon [Wheeler (1955); Brill and Hartle (1964); for more on gravitational-wave energy, see §35.14] an object built out of gravitational waves (or electromagnetic waves, or neutrinos, or any combination of the three) that holds itself together for a time that is long in comparison to the characteristic period of vibration of the waves. It looks from a distance like any other mass, even though nowhere in its interior can one put a finger and say "here is mass." Therefore it, like any other mass, must have "its influence on inertia." But to specify this mass, one must give enough information to characterize completely the gravitational waves on the simultaneity Σ. For this, it is not enough merely to have given the two "wave-coordinates" per spacepoint that one possesses in ${}^{(3)}\mathcal{G}$. One must give in addition (7) the two "wave-momenta" per spacepoint that appear in York's "momentum density of weight 5/3," $\tilde{\pi}^{ab}$; and at the same time, as an inextricable part of this operation, one must (8) specify the density of flow of field energy. (9) Solve for the conformal factor ψ. (10) Then one has complete initial-value data that satisfy the initial-value equations of general relativity. (11) These data now known, the remaining, dynamic, components of the field equation determine the 4-geometry into the past and the future. (12) In this way, the inertial properties of every test particle are determined everywhere and at all times, giving concrete realization to Mach's principle.

Much must still be done to spell out the physics behind these equations and to see this physics in action. Some significant progress had already been made in this direction before the present stage in one's understanding of the initial-value equations. Especially interesting are results of Thirring (1918) and (1921) and of Thirring and Lense (1918), discussed by Einstein (1950) in the third edition of his book, *The Meaning of Relativity.*

The Foucault pendulum

Consider a bit of solid ground near the geographic pole, and a support erected there, and from it hanging a pendulum. Though the sky is cloudy, the observer watches the track of the Foucault pendulum as it slowly turns through 360°. Then the sky clears and, miracle of miracles, the pendulum is found to be swinging all the time on an arc fixed relative to the far-away stars. If "mass there governs inertia here," as envisaged by Mach, how can this be?

The dragging of the inertial frame

Enlarge the question. By the democratic principle that equal masses are created equal, the mass of the earth must come into the bookkeeping of the Foucault pendulum. Its plane of rotation must be dragged around with a slight angular velocity, ω_{drag}, relative to the so-called "fixed stars." How much is ω_{drag}? And how much would ω_{drag} be if the pendulum were surrounded by a rapidly spinning spherical shell of mass M and radius R_{shell}, turning at angular velocity ω_{shell}?

Einstein's theory says that inertia is a manifestation of the geometry of spacetime. It also says that geometry is affected by the presence of matter to an extent proportional to the factor $G/c^2 = 0.742 \times 10^{-28}$ cm/g. Simple dimensional considerations leave no room except to say that the rate of drag is proportional to a expression of the form

$$\omega_{\text{drag}} = k\frac{G}{c^2}\frac{m_{\text{shell, conv}}}{R_{\text{shell}}}\omega_{\text{shell}} = k\frac{m_{\text{shell}}}{R_{\text{shell}}}\omega_{\text{shell}}. \tag{21.155}$$

Here k is a numerical factor to be found only by detailed calculation. Lense and Thirring [(1918) and (1921)], starting with a flat background spacetime manifold, calculated in the weak-field approximation of Chapter 18 the effect of the moving current of mass on the metric. Expressed in polar coordinates, the metric acquires a non-zero coefficient $g_{\phi t}$. Inserted into the equation of geodesic motion, this off-diagonal metric coefficient gives rise to a precession. This precession (defined here about an axis parallel to the axis of rotation, not about the local vertical) is given by an expression of the form (21.155), where the precession factor k has the value

$$k = 4/3. \tag{21.156}$$

There is a close parallelism between the magnetic component of the Maxwell field and the precession component of the Einstein field. In neither field does a source at rest produce the new kind of effect when acting on a test particle that is also at rest. One designs a circular current of charge to produce a magnetic field; and a test charge, in order to respond to this magnetic field, must also be in motion. Similarly here: no pendulum vibration means no pendulum precession. Moreover, the direction of the precession depends on where the pendulum is, relative to the rotating shell of mass. The precession factor k has the following values:

$$\begin{aligned} k &= 4/3 \quad \text{for pendulum anywhere inside rotating shell of mass;} \\ k &= 4/3 \quad \text{for pendulum at North or South pole;} \\ k &= -2/3 \quad \text{for pendulum just outside the rotating shell at its equator.} \end{aligned} \tag{21.157}$$

This position-dependence of the drag, ω_{drag}, makes still more apparent the analogy with magnetism, where the field of a rotating charged sphere points North at the center of the sphere, and North at both poles, but South at the equator.

Whether the Foucault pendulum is located in imagination at the center of the earth or in actuality at the North pole, the order of magnitude of the expected drag is

$$\begin{aligned} \omega_{\text{drag}} &\sim \frac{m_{\text{earth}}}{R_{\text{earth}}}\omega_{\text{earth}} \sim \frac{0.44 \text{ cm}}{6 \times 10^8 \text{ cm}} \frac{1 \text{ radian}}{13700 \text{ sec}} \\ &\sim 5 \times 10^{-14} \text{ rad/sec,} \end{aligned} \tag{21.158}$$

too small to allow detection, let alone actual measurement, by any device so far built—but perhaps measurable by gyroscopes now under construction (§40.7). By contrast, near a rapidly spinning neutron star or near a black hole endowed with substantial angular momentum, the calculated drag effect is not merely detectable; it is even important (see Chapter 33 on the physics of a rotating black hole).

The distant stars must influence the natural plane of vibration of the Foucault pendulum as the nearby rotating shell of matter does, provided that the stars are not so far away ($r \sim$ radius of universe) that the curvature of space begins to introduce substantial corrections into the calculation of Thirring and Lense. In other words, no reason is apparent why all masses should not be treated on the same footing, so that (21.158) more appropriately, if also somewhat symbolically, reads

The "sum for inertia"

$$\omega_{\substack{\text{plane of} \\ \text{vibration} \\ \text{of Foucault} \\ \text{pendulum}}} \sim \frac{m_{\text{shell}}}{R_{\text{shell}}}\omega_{\text{shell}} + \sum_{\substack{\text{far-away} \\ \text{"stars"}}} \frac{m_{\text{"star"}}}{r_{\text{"star"}}}\omega_{\text{"star"}}. \tag{21.159}$$

Moreover, when there is no nearby shell of matter, or when it has negligible effects, the plane of vibration of the pendulum, if experience is any guide, cannot turn with respect to the frame defined by the far-away "stars." In this event ω_{Foucault} must be identical with ω_{stars}; or the "sum for inertia,"

$$\sum_{\substack{\text{far-away} \\ \text{"stars"}}} \frac{m_{\text{"star"}}}{r_{\text{"star"}}} \sim \frac{m_{\text{universe}}}{r_{\text{universe}}}, \tag{21.160}$$

must be of the order of unity. Just such a relation of approximate identity between the mass content of the universe and its radius at the phase of maximum expansion is a characteristic feature of the Friedman model and other simple models of a closed universe (Chapters 27 and 30). In this respect, Einstein's theory of Mach's principle exhibits a satisfying degree of self-consistency.

At phases of the dynamics of the universe other than the stage of maximum expansion, r_{universe} can become arbitrarily small compared to m_{universe}. Then the ratio (21.160) can depart by powers of ten from unity. Regardless of this circumstance, one has no option but to understand that the *effective* value of the "sum for inertia" is still unity after all corrections have been made for the dynamics of contraction or expansion, for retardation, etc. Only so can ω_{Foucault} retain its inescapable identity with $\omega_{\text{far-away stars}}$. Fortunately, one does not have to pursue the theology of the "sum for inertia" to the uttermost of these sophistications to have a proper account of inertia. Mach's idea that mass there determines inertia here has its complete mathematical account in Einstein's geometrodynamic law, as already spelled out. For the first strong-field analysis of the dragging of the inertial reference system in the context of relativistic cosmology, see Brill and Cohen (1966) and Cohen and Brill (1967); see also §33.4 for dragging by a rotating black hole.

Minkowski geometry as limit of a closed 3-geometry

Still another clarification is required of what Mach's principle means and how it is used. The inertial properties of a test particle are perfectly well-determined when that particle is moving in ideal Minkowski space. "Point out, please," the anti-Machian critic says, "the masses that are responsible for this inertia." In answer, recall that Einstein's theory includes not only the geometrodynamic law, but also, in Einstein's view, the boundary condition that the universe be closed. Thus the section of spacetime that is flat is to be viewed, not as infinite, but as part of a closed universe. (For a two-dimensional analog, fill a rubber balloon with water and set it on a glass tabletop and look at it from underneath). The part of the universe that is curved acquires its curvature by reason of its actual content of mass-energy or—if animated only by gravitational waves—by reason of its effective content of mass-energy. This mass-energy, real or effective, is to be viewed as responsible for the inertial properties of the test particle that at first sight looked all alone in the universe.

It in no way changes the qualitative character of the result to turn attention to a model universe where the region of Minkowski flatness, and all the other linear dimensions of the universe, have been augmented tenfold ("ten times larger balloon; ten times larger face"). The curvature and density of the curved part of the model universe are down by a factor of 100, the volume is up by a factor of 1,000, the mass is up by a factor of 10; but the ratio of mass to radius, or the "sum for inertia" (the poor man's substitute for a complete initial-value calculation) is unchanged.

Einstein acknowledged a debt of parentage for his theory to Mach's principle (Figure 21.5). It is therefore only justice that Mach's principle should in return today owe its elucidation to Einstein's theory.

EXERCISES

Exercise 21.21. WHY THE WEYL CONFORMAL CURVATURE TENSOR VANISHES

How many independent components does the Riemann curvature tensor have in three-dimensional space? How many does the Ricci curvature tensor have? Show that the two tensors are related by the formula

$$R^d{}_{abc} = \delta^d_b R_{ac} - \delta^d_c R_{ab} + g_{ac} R^d{}_b - g_{ab} R^d_c + \frac{1}{2} R(\delta^d_c g_{ab} - \delta^d_b g_{ac})$$

with no need of any Weyl conformal-curvature tensor to specify (as in higher dimensions) the further details of the Riemann tensor. Show that the Weyl tensor, from an n-dimensional modification of equation (13.50) as in exercise 13.13, vanishes for $n = 2$.

Exercise 21.22. YORK'S CURVATURE

[York (1971)]. (a) Define the tensor [Eisenhart (1926)]

$$R_{abc} = R_{ab|c} - R_{ac|b} + \frac{1}{4}(g_{ac} R_{|b} - g_{ab} R_{|c}).$$

(b) Show that a 3-geometry is conformally flat when and only when $R_{abc} = 0$.

(c) Show that the following identities hold and reduce to five the number of independent components of R_{abc}:

$$R^a{}_{ac} = g^{ab} R_{bac} = 0;$$
$$R_{abc} + R_{acb} = 0;$$
$$R_{abc} + R_{cab} + R_{bca} = 0.$$

(d) Show that Yorks' curvature

$$\begin{aligned} Y^{ab} &= g^{1/3}[aef]\left(R_f{}^b - \frac{1}{4}\delta_f{}^b R\right)_{|e} \\ &= -\frac{1}{2} g^{1/3}[aef] g^{bm} R_{mef} \end{aligned}$$

is conformally invariant and has the properties listed in equations (21.148).

Exercise 21.23. PULLING THE POYNTING FLUX VECTOR "OUT OF THE AIR"

From the condition that the Hamilton-Jacobi functional $S(g_{ij}, A_m)$ (extremal of the action integral) for the combined Einstein and Maxwell fields, ostensibly dependent on the six metric coefficients $g_{ij}(x, y, z)$ and the three potentials $A_m(x, y, z)$, shall actually depend only on the 3-geometry of the spacelike hypersurface and the distribution of magnetic field strength on this hypersurface, show that the geometrodynamic field momentum $\pi^{ij} = \delta S/\delta g_{ij}$ satisfies a condition of the form

$$\pi^{ij}{}_{|j} = c[imn]\mathcal{E}_m \mathcal{B}_n,$$

and evaluate the coefficient c in this equation [Wheeler (1968b)]. Hint: Note that the transformation

$$x^i \longrightarrow x^i - \xi^i,\ g_{ij} \longrightarrow g_{ij} + \xi_{i|j} + \xi_{j|i}$$

in no way changes the 3-geometry itself, and therefore the corresponding induced change in S,

$$\delta S = \int \left[\frac{\delta S}{\delta g_{ij}}\delta g_{ij} + \frac{\delta S}{\delta A_m}\delta A_m\right] d^3x$$

must vanish identically for arbitrary choice of the $\xi^i(x, y, z)$, which measure the equivalent of the sliding of a ruled transparent rubber sheet over an automobile fender.

Exercise 21.24. THE EXTREMAL ACTION ASSOCIATED WITH THE HILBERT ACTION PRINCIPLE DEPENDS ON CONFORMAL 3-GEOMETRY AND EXTRINSIC TIME [K. Kuchař (1972) and J. York (1972)]

Show that the data demanded by the Hilbert action principle $\delta \int {}^{(4)}R(-{}^{(4)}g)^{1/2}\, d^4x = 0$ on each of the two bounding spacelike hypersurfaces consist of (1) the conformal 3-geometry ${}^{(3)}<$ of the hypersurface plus (2) the extrinsic time variable defined by

$$\tau = \frac{2}{3} g^{-1/2}\, \mathrm{Tr}\, \boldsymbol{\pi} = \frac{4}{3}\, \mathrm{Tr}\, \boldsymbol{K},$$

conveniently represented by the pictogram $\succ\!\!\prec$, measured by one number per spacepoint, and independent of the conformal factor in the metric of the 3-geometry. This done, explain in a few words why in this formulation of geometrodynamics the Hamilton-Jacobi function ($\hbar$ times the phase of the wave function in the semiclassical or JWKB approximation) is appropriately expressed in the form

$$S = S({}^{(3)}<, \succ\!\!\prec).$$

§21.13. JUNCTION CONDITIONS

The intrinsic and extrinsic curvatures of a hypersurface, which played such fundamental roles in the initial-value formalism, are also powerful tools in the analysis of "junction conditions."

Junction conditions for electrodynamics

Recall the junction conditions of electrodynamics: across any surface (e.g., a capacitor plate), the tangential part of the electric field, $\boldsymbol{E}_{\parallel}$, and the normal part of the magnetic field, $\boldsymbol{B}_{\perp}$, must be continuous; thus,

$$\begin{aligned}
[\boldsymbol{E}_{\parallel}] &\equiv (\text{discontinuity in } \boldsymbol{E}_{\parallel}) \\
&\equiv (\boldsymbol{E}_{\parallel} \text{ on "+" side of surface}) - (\boldsymbol{E}_{\parallel} \text{ on "−" side of surface}) \\
&\equiv \boldsymbol{E}_{\parallel}^{+} - \boldsymbol{E}_{\parallel}^{-} = 0, \qquad (21.161\text{a})\\
[\boldsymbol{B}_{\perp}] &\equiv \boldsymbol{B}_{\perp}^{+} - \boldsymbol{B}_{\perp}^{-} = 0; \qquad (21.161\text{b})
\end{aligned}$$

while the "jump" in the parts $\boldsymbol{E}_{\perp}$ and $\boldsymbol{B}_{\parallel}$ must be related to the charge density (charge per unit area) σ, the current density (current per unit area) $\boldsymbol{j}$, and the unit normal to the surface $\boldsymbol{n}$ by the formulas

$$[\boldsymbol{E}_{\perp}] = \boldsymbol{E}_{\perp}^{+} - \boldsymbol{E}_{\perp}^{-} = 4\pi\sigma\boldsymbol{n}, \qquad (21.161\text{c})$$

$$[\boldsymbol{B}_{\parallel}] = \boldsymbol{B}_{\parallel}^{+} - \boldsymbol{B}_{\parallel}^{-} = 4\pi\boldsymbol{j} \times \boldsymbol{n}. \qquad (21.161\text{d})$$

Recall also that one derives these junction conditions by integrating Maxwell's equations over a "pill box" that is centered on the surface.

Similar junction conditions, derivable in a similar manner, apply to the gravitational field (spacetime curvature), and to the stress-energy that generates it.* Focus

*The original formulation of gravitational junction conditions stemmed from Lanczos (1922, 1924). The formulation given here, in terms of intrinsic and extrinsic curvature, was developed by Darmois (1927), Misner and Sharp (1964), and Israel (1966). For further references to the extensive literature, see Israel.

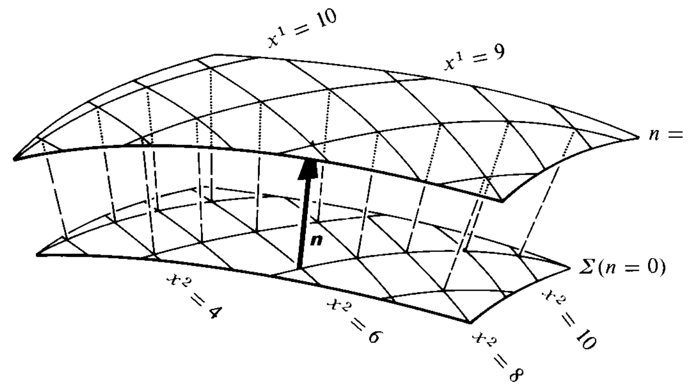


Figure 21.6.
Gaussian normal coordinates in the neighborhood of a 3-surface Σ. The metric in Gaussian normal coordinates has the form

$$ds^2 = (\boldsymbol{n}\cdot\boldsymbol{n})^{-1}\, dn^2 + g_{ij}\, dx^i\, dx^j$$

with $\boldsymbol{n} = \partial/\partial n$, $(\boldsymbol{n}\cdot\boldsymbol{n}) = -1$ if the surface is spacelike, and $(\boldsymbol{n}\cdot\boldsymbol{n}) = 1$ if it is timelike. (See exercise 27.2.) The extrinsic curvature of the surfaces $n =$ constant is $K_{ij} = -\frac{1}{2}\,\partial g_{ij}/\partial n$, and the Einstein field equations written in "3 + 1" form are (21.162).

attention on a specific three-dimensional slice through spacetime—the 3-surface Σ on Figure 21.6. Let the surface be either spacelike [unit normal $\boldsymbol{n}$ timelike; $(\boldsymbol{n}\cdot\boldsymbol{n}) = -1$] or timelike [$\boldsymbol{n}$ spacelike; $(\boldsymbol{n}\cdot\boldsymbol{n}) = +1$]. The null case will be discussed later. As an aid in deriving junction conditions, introduce Gaussian normal coordinates in the neighborhood of Σ [see the paragraph preceeding equation (21.82)]. In terms of the intrinsic and extrinsic curvatures of Σ and of neighboring 3-surfaces $n =$ constant, the Einstein tensor and Einstein field equation have components

Einstein equation in "3 + 1" form

$$G^n{}_n = -\frac{1}{2}{}^{(3)}R + \frac{1}{2}(\boldsymbol{n}\cdot\boldsymbol{n})^{-1}\{(\mathrm{Tr}\,\boldsymbol{K})^2 - \mathrm{Tr}\,(\boldsymbol{K}^2)\} = 8\pi T^n{}_n, \tag{21.162a}$$

$$G^n{}_i = -(\boldsymbol{n}\cdot\boldsymbol{n})^{-1}\{K_i{}^m{}_{|m} - (\mathrm{Tr}\,\boldsymbol{K})_{|i}\} = 8\pi T^n{}_i, \tag{21.162b}$$

$$\begin{aligned} G^i{}_j = {}^{(3)}G^i{}_j + (\boldsymbol{n}\cdot\boldsymbol{n})^{-1}\{(K^i{}_j - \delta^i{}_j\,\mathrm{Tr}\,\boldsymbol{K})_{,n} \\ - (\mathrm{Tr}\,\boldsymbol{K})K^i{}_j + \frac{1}{2}\delta^i{}_j(\mathrm{Tr}\,\boldsymbol{K})^2 + \frac{1}{2}\delta^i{}_j\,\mathrm{Tr}\,(\boldsymbol{K}^2)\} = 8\pi T^i{}_j. \end{aligned} \tag{21.162c}$$

[See equations (21.77), (21.81), (21.76), and (21.82).]

Suppose that the stress-energy tensor $T^\alpha{}_\beta$ contains a "delta-function singularity" at Σ—i.e., suppose that Σ is the "world tube" of a two-dimensional surface with finite 4-momentum per unit area (analog of surface charge and surface current in electrodynamics). Then define the *surface stress-energy tensor on Σ* to be the integral of $T^\alpha{}_\beta$ with respect to proper distance (n), measured perpendicularly through Σ:

Surface stress-energy tensor

$$S^\alpha{}_\beta = \lim_{\varepsilon\to 0}\left[\int_{-\varepsilon}^{+\varepsilon} T^\alpha{}_\beta\, dn\right]. \tag{21.163}$$

To discover the effect of this surface layer on the spacetime geometry, perform a "pill-box integration" of the Einstein field equation (21.162)

Derivation of junction conditions

$$\lim_{\varepsilon\to 0}\left[\int_{-\varepsilon}^{+\varepsilon} G^{\alpha}{}_{\beta}\, dn\right] = 8\pi S^{\alpha}{}_{\beta}. \tag{21.164}$$

Examine the integral of $G^{\alpha}{}_{\beta}$. If the 3-metric g_{ij} were to contain a delta function or a discontinuity at Σ, then Σ would not have any well-defined 3-geometry—a physically inadmissible situation, even in the presence of surface layers. Absence of delta functions, $\delta(n)$, in g_{ij} means absence of delta functions in ${}^{(3)}R$; absence of discontinuities in g_{ij} means absence of delta functions in $K_{ij} = -\frac{1}{2}g_{ij,n}$. Thus, equations (21.162) when integrated say

$$\int G^{n}{}_{n}\, dn = 0 = 8\pi S^{n}{}_{n}, \tag{21.165a}$$

$$\int G^{n}{}_{i}\, dn = 0 = 8\pi S^{n}{}_{i}, \tag{21.165b}$$

$$\int G^{i}{}_{j}\cdot dn = (\boldsymbol{n}\cdot\boldsymbol{n})(\gamma^{i}{}_{j} - \delta^{i}{}_{j}\,\mathrm{Tr}\,\boldsymbol{\gamma}) = 8\pi S^{i}{}_{j}, \tag{21.165c}$$

where $\gamma^{i}{}_{j}$ is the "jump" in the components of the extrinsic curvature

$$\begin{aligned}\boldsymbol{\gamma} \equiv [\boldsymbol{K}] &\equiv (\boldsymbol{K} \text{ on "}n = +\varepsilon \text{ side" of } \Sigma) - (\boldsymbol{K} \text{ on "}n = -\varepsilon \text{ side" of } \Sigma)\\ &\equiv \boldsymbol{K}^{+} - \boldsymbol{K}^{-}.\end{aligned} \tag{21.166}$$

Junction conditions in absence of surface layers

In the absence of a delta-function surface layer, the above junction conditions say, simply, that $\boldsymbol{\gamma} \equiv [\boldsymbol{K}] = 0$. In words: if one examines how Σ is embedded in the spacetime above its "upper"face, and how it is embedded in the spacetime below its "lower" face, one must discover identical embeddings—i.e., identical extrinsic curvatures $\boldsymbol{K}$. Of course, the intrinsic curvature of Σ must also be the same, whether viewed from above or below. More briefly:

$$(\text{absence of surface layers}) \rightleftarrows (\text{"continuity" of } g_{ij} \text{ and } K_{ij}). \tag{21.167}$$

Junction conditions for a surface layer

If a surface layer is present, then Σ must be the world tube of a two-dimensional layer of matter, and the normal to Σ must be spacelike, $(\boldsymbol{n}\cdot\boldsymbol{n}) = +1$. The junction conditions (21.165a,b) then have the simple physical meaning

$$\boldsymbol{S}(\boldsymbol{n},\ldots) = 0 \leftrightarrows \begin{pmatrix}\text{the momentum flow is entirely in } \Sigma;\\ \text{i.e., no momentum associated with the}\\ \text{surface layer flows out of } \Sigma;\text{ i.e., } \Sigma\\ \text{is the world tube of the surface layer}\end{pmatrix}, \tag{21.168a}$$

which tells one nothing new. The junction condition (21.165c) says that the surface stress-energy generates a discontinuity in the extrinsic curvature (different embedding in spacetime "above" Σ than "below" Σ), given by

$$\gamma^{i}{}_{j} - \delta^{i}{}_{j}\,\mathrm{Tr}\,\boldsymbol{\gamma} = 8\pi S^{i}{}_{j}. \tag{21.168b}$$

Of course, the intrinsic geometry of Σ must be the same as seen from above and below,

$$g_{ij} \text{ continuous across } \Sigma. \tag{21.169}$$

2

In analyzing surface layers, one uses not only the junction conditions (21.168a) to (21.169), but also the four-dimensional Einstein field equation applied on each side of the surface Σ separately, and also an equation of motion for the surface stress-energy. The equation of motion is derived by examining the jump in the field equation $G^n{}_i = 8\pi T^n{}_i$ (equation 21.162b); thus $[G^n{}_i] = 8\pi[T^n{}_i]$ says

$$(\gamma_i{}^m - \delta_i{}^m \operatorname{Tr} \boldsymbol{\gamma})_{|m} = -8\pi[T^n{}_i];$$

and when reexpressed in terms of $S_i{}^m$ by means of the junction condition (21.168b), it says

Equation of motion for a surface layer

$$S^{im}{}_{|m} + [T^{in}] = 0. \tag{21.170}$$

[For intuition into this equation of motion, see Exercises 21.25 and 21.26. For applications of the "surface-layer formalism" see exercise 21.27; also Israel (1966), Kuchař (1968), Papapetrou and Hamoui (1968).]

Gravitational-wave shock fronts

When one turns attention to junction conditions across a *null* surface Σ, one finds results rather different from those in the spacelike and timelike cases. A "pill-box" integration of the field equations reveals that even in vacuum the extrinsic curvature may be discontinuous. A discontinuity in K_{ij} across a null surface, without any stress-energy to produce it, is the geometric manifestation of a *gravitational-wave shock front* (analog of a shock-front in hydrodynamics). For quantitative details see, e.g., Pirani (1957), Papapetrou and Treder (1959, 1962), Treder (1962), and especially Choquet-Bruhat (1968b).

That a discontinuity in the curvature tensor can propagate with the speed of light is a reminder that all gravitational effects, like all electromagnetic effects, obey a causal law. The initial-value data on a spacelike initial-value hypersurface uniquely determine the resulting spacetime geometry [see the work of Cartan, Stellmacher, Lichnerowicz, and Bruhat (also under the names Fourès-Bruhat and Choquet-Bruhat) and others cited and summarized in the article of Bruhat (1962)] but determine it in a way consistent with causality. Thus a change in these data throughout a limited region of the initial value 3-geometry makes itself felt on a slightly later hypersurface solely in a region that is also limited, and only a little larger than the original region.

When one turns from classical dynamics to quantum dynamics, one sees new reason to focus attention on a spacelike initial-value hypersurface: the observables at different points on such a hypersurface commute with one another; i.e., are in principle simultaneously observable.

Not every four-dimensional manifold admits a global singularity-free spacelike hypersurface. Those manifolds that do admit such a hypersurface have more to do with physics, it is possible to believe, than those that do not.

Even in a manifold that does admit a spacelike hypersurface, attention has been given sometimes, in the context of classical theory, to initial-value data on a hypersurface that is not spacelike but "characteristic," in the sense that it accomodates null geodesics [see, for example, Sachs (1964) and references cited there]. It is typical in such situations that one can predict the future but not the past, or predict the past but not the future.

Children of light and children of darkness is the vision of physics that emerges from this chapter, as from other branches of physics. The children of light are the differential equations that predict the future from the present. The children of darkness are the factors that fix these initial conditions.

EXERCISES

Exercise 21.25. EQUATION OF MOTION FOR A SURFACE LAYER

(a) Let $\boldsymbol{u}$ be the "mean 4-velocity" of the matter in a surface layer—so defined that an observer moving with 4-velocity $\boldsymbol{u}$ sees zero energy flux. Let σ be the total mass-energy per unit proper surface area, as measured by such a "comoving observer." Show that the surface stress-energy tensor can be expressed in the form

$$\boldsymbol{S} = \sigma \boldsymbol{u} \otimes \boldsymbol{u} + \boldsymbol{t}, \text{ where } (\boldsymbol{t} \cdot \boldsymbol{u}) = 0, \tag{21.171}$$

and where $\boldsymbol{t}$ is a symmetric stress tensor.

(b) Show that the component along $\boldsymbol{u}$ of the equation of motion (21.170) is

$$d\sigma/d\tau = -\sigma u^j{}_{|j} + u_j t^{jk}{}_{|k} + u_j[T^{jn}], \tag{21.172}$$

where $d/d\tau = \boldsymbol{u}$. Give a physical interpretation for each term.

(c) Let a_j be that part of the 4-acceleration of the comoving observer which lies in the surface layer Σ. By projecting the equation of motion (21.170) perpendicular to $\boldsymbol{u}$, show that

$$\sigma a_j = -P_{ja}\{t^{ab}{}_{|b} + [T^{an}]\}, \tag{21.173}$$

where P_{ja} is the projection operator

$$P_{ja} = g_{ja} + u_j u_a. \tag{21.174}$$

Give a physical interpretation for each term of equation (21.182).

Exercise 21.26. THIN SHELLS OF DUST

For a thin shell of dust surrounded by vacuum ($[T^{jn}] = 0$, $\boldsymbol{t} = 0$), derive the following equations

$$d\sigma/d\tau = -\sigma u^b{}_{|b}, \tag{21.175a}$$

$$\boldsymbol{a}^+ + \boldsymbol{a}^- = 0, \tag{21.175b}$$

$$\boldsymbol{a}^+ - \boldsymbol{a}^- = (4\pi\sigma)\boldsymbol{n} \tag{21.175c}$$

$$\boldsymbol{\gamma} = 8\pi\sigma\left(\boldsymbol{u} \otimes \boldsymbol{u} + \frac{1}{2}\boldsymbol{g}\right). \tag{21.175d}$$

Here $\boldsymbol{a}^+$ and $\boldsymbol{a}^-$ are the 4-accelerations as measured by accelerometers that are fastened onto the outer and inner sides of the shell, and $\boldsymbol{g}$ is the 3-metric of the shell. Show that the first of these equations is the law of "conservation of rest mass."

Exercise 21.27. SPHERICAL SHELL OF DUST

Apply the formalism of exercise 21.25 to a collapsing spherical shell of dust [Israel (1967b)]. For the metric inside and outside the shell, take the flat-spacetime and vacuum Schwarzschild expressions (Chapter 23),

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \text{ inside,} \tag{21.176a}$$

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{1 - 2M/r} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \text{ outside.} \tag{21.176b}$$

Let the "radius" of the shell, as a function of proper time measured on the shell, be

$$R \equiv \frac{1}{2\pi} \times (\text{proper circumference of shell}) = R(\tau). \tag{21.176c}$$

Show that the shell's mass density varies with time as

$$\sigma(\tau) = \mu/4\pi R^2(\tau), \qquad \mu = \text{constant} = \text{"total rest mass"}; \tag{21.176d}$$

and derive and solve the equation of motion

$$M = \mu\left\{1 + \left(\frac{dR}{d\tau}\right)^2\right\}^{1/2} - \frac{\mu}{2R}. \tag{21.176e}$$

CHAPTER 22

THERMODYNAMICS, HYDRODYNAMICS, ELECTRODYNAMICS, GEOMETRIC OPTICS, AND KINETIC THEORY

§22.1. THE WHY OF THIS CHAPTER

This chapter is entirely Track 2. No earlier Track-2 material is needed as preparation for it, but Chapter 5 (stress-energy tensor) will be helpful.

§22.5 (geometric optics) is needed as preparation for Chapter 34 (singularities and global methods). The rest of the chapter is not needed as preparation for any later chapter; but it will be extremely helpful in most applications of gravitation theory (Chapters 23–40).

Astrophysical applications of gravitation theory are the focus of the rest of this book, except for Chapters 41–44. Each application—stars, star clusters, cosmology, collapse, black holes, gravitational waves, solar-system experiments—can be pursued by itself at an elementary level, without reference to the material in this chapter. But deep understanding of the applications requires a prior grasp of thermodynamics, hydrodynamics, electrodynamics, geometric optics, and kinetic theory, all in the context of curved spacetime. Hence, most Track-2 readers will want to probe these subjects at this point.

§22.2. THERMODYNAMICS IN CURVED SPACETIME*

Thermodynamic potentials are defined in rest frame of fluid

Consider, for concreteness and simplicity, the equilibrium thermodynamics of a perfect fluid with fixed chemical composition ("simple perfect fluid")—for example, the gaseous interior of a collapsing supermassive star. The thermodynamic state of a fluid element, as it passes through an event $\mathscr{P}_0$, can be characterized by various thermodynamic potentials, such as n, ρ, p, T, s, μ. The numerical value of each potential at $\mathscr{P}_0$ is measured in the proper reference frame (§13.6) of an observer who moves with the fluid element—i.e., in the fluid element's "rest frame." Despite

*For more detailed treatments of this subject see, e.g., Stueckelberg and Wanders (1953), Kluitenberg and de Groot (1954), Meixner and Reik (1959), and references cited therein; see also the references on hydrodynamics cited at the beginning of §22.3, and the references on kinetic theory cited at the beginning of §22.6.

this use of rest frame to measure the potentials, the potentials are frame-independent functions (scalar fields). At the chosen event $\mathscr{P}_0$, a given potential (e.g., n) has a unique value $n(\mathscr{P}_0)$; so n is a perfectly good frame-independent function.

The values of n, ρ, p, T, s, μ measure the following quantities in the rest frame of the fluid element:

Definitions of thermodynamic potentials

n, baryon number density; i.e., number of baryons per unit three-dimensional volume of rest frame, with antibaryons (if any) counted negatively.
ρ, density of total mass-energy; i.e., total mass-energy (including rest mass, thermal energy, compressional energy, etc.) contained in a unit three-dimensional volume of the rest frame.
p, isotropic pressure in rest frame.
T, temperature in rest frame.
s, entropy per baryon in rest frame. (The entropy per unit volume is ns.)
μ, chemical potential of baryons in rest frame [see equation (22.8) below].

Definition of "simple fluid"

The chemical composition of the fluid (number density of hydrogen molecules, number density of hydrogen atoms, number density of free protons and electrons, number density of photons, number density of ^{238}U nuclei, number density of Λ hyperons . . .) is assumed to be fixed uniquely by two thermodynamic variables—e.g., by the total number density of baryons n and the entropy per baryon s. In this sense the fluid is a "simple fluid." Simple fluids occur whenever the chemical abundances are "frozen" (reaction rates too slow to be important on the time scales of interest; for example, in a supermassive star except during explosive burning and except at temperatures high enough for $e^- - e^+$ pair production). Simple fluids also occur in the opposite extreme of complete chemical equilibrium (reaction rates fast enough to maintain equilibrium despite changing density and entropy; for example, in neutron stars, where high pressures speed up all reactions). When one examines nuclear burning in a nonconvecting star, or explosive nuclear burning, or pair production and neutrino energy losses at high temperatures, one must usually treat the fluid as "multicomponent." Then one introduces a number density n_J and a chemical potential μ_J for each chemical species with abundance not fixed by n and s. For further details see, e.g., Zel'dovich and Novikov (1971).

Law of baryon conservation

The most fundamental law of thermodynamics—even more fundamental than the "first" and "second" laws—is *baryon conservation*. Consider a fluid element whose moving walls are attached to the fluid so that no baryons flow in or out. As the fluid element moves through spacetime, deforming along the way, its volume V changes. But the number of baryons in it must remain fixed, so

$$\frac{d}{d\tau}(nV) = 0. \tag{22.1}$$

The changes in volume are produced by the flow of neighboring bits of fluid away from or toward each other—explicitly (exercise 22.1)

$$dV/d\tau = (\boldsymbol{\nabla} \cdot \boldsymbol{u})V, \tag{22.2}$$

where $\boldsymbol{u} = d/d\tau$ is the 4-velocity of the fluid. Consequently, baryon conservation [equation (22.1)] can be reexpressed as

$$0 = \frac{dn}{d\tau} + \frac{n}{V}\frac{dV}{d\tau} = \boldsymbol{\nabla}_{\boldsymbol{u}} n + n(\boldsymbol{\nabla} \cdot \boldsymbol{u}) = \boldsymbol{u} \cdot \boldsymbol{\nabla} n + n(\boldsymbol{\nabla} \cdot \boldsymbol{u}) = \boldsymbol{\nabla} \cdot (n\boldsymbol{u});$$

i.e.,

$$\boldsymbol{\nabla} \cdot \boldsymbol{S} = 0, \tag{22.3}$$

$$\boldsymbol{S} = n\boldsymbol{u} = \text{baryon number-flux vector} \tag{22.4}$$

(see §5.4 and exercise 5.3.) Moreover, this abstract geometric version of the law must be just as valid in curved spacetime as in flat (equivalence principle).

Note the analogy with the law of charge conservation, $\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$, in electrodynamics (exercise 3.16) and with the local law of energy-momentum conservation, $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ (§§5.9 and 16.2). In a very deep sense, the forms of these three laws are dictated by the theorem of Gauss (§5.9, and Boxes 5.3, 5.4).

Second law of thermodynamics

The second law of thermodynamics states that, in flat spacetime or in curved, entropy can be generated but not destroyed. Apply this law to a fluid element of volume V containing a fixed number of baryons N. The entropy it contains is

$$S = Ns = nsV.$$

Entropy may flow in and out across the faces of the fluid element ("heat flow" between neighboring fluid elements); but for simplicity assume it does not; or if it does, assume that it flows too slowly to have any significance for the problem at hand. Then the entropy in the fluid element can only increase:

$$d(nsV)/d\tau \geq 0 \quad \text{when negligible entropy is exchanged between neighboring fluid elements;}$$

i.e. [combine with equation (22.1)]

$$ds/d\tau \geq 0 \text{ (no entropy exchange).} \tag{22.5}$$

Shock waves and heat flow

So long as the fluid element remains in thermodynamic equilibrium, its entropy will actually be conserved ["$=$" in equation (22.5)]; but at a shock wave, where equilibrium is momentarily broken, the entropy will increase (conversion of "relative kinetic energy" of neighboring fluid elements into heat). [For discussions of heat flow in special and general relativity, see Exercise 22.7. For discussion of shock waves, see Taub (1948), de Hoffman and Teller (1950), Israel (1960), May and White (1967), Zel'dovich and Rayzer (1967), Lichnerowicz (1967, 1971), and Thorne (1973a).]

First law of thermodynamics

The first law of thermodynamics, in the proper reference frame of a fluid element, is identical to the first law in flat spacetime ("principle of equivalence"); and in flat spacetime the first law is merely the law of energy conservation:

$$d\begin{pmatrix}\text{energy in a volume element containing}\\ \text{a fixed number, } A\text{, of baryons}\end{pmatrix} = -p\, d(\text{volume}) + T\, d(\text{entropy});$$

i.e.,

$$d(\rho A/n) = -p\, d(A/n) + T\, d(As);$$

i.e.,

$$d\rho = \frac{\rho + p}{n}\, dn + nT\, ds.$$

Query: what kind of a "d" appears here? For a simple fluid, the values of two potentials, e.g., n and s, fix all the others uniquely; so *any* change in ρ must be determined uniquely by the changes in n and s. It matters not whether the changes are measured along the world line of a given fluid element, or in some other direction. Thus, the "d" in the first law can be interpreted as an exterior derivative

$$\boldsymbol{d}\rho = \frac{\rho + p}{n}\, \boldsymbol{d}n + nT\, \boldsymbol{d}s; \tag{22.6}$$

and the changes along a given direction in the fluid (along a given tangent vector $\boldsymbol{v}$) can be written

$$\begin{aligned}\boldsymbol{\nabla}_{\boldsymbol{v}}\rho \equiv \langle \boldsymbol{d}\rho, \boldsymbol{v}\rangle &= \frac{\rho + p}{n} \langle \boldsymbol{d}n, \boldsymbol{v}\rangle + nT\langle \boldsymbol{d}s, \boldsymbol{v}\rangle \\ &= \frac{\rho + p}{n} \boldsymbol{\nabla}_{\boldsymbol{v}} n + nT\, \boldsymbol{\nabla}_{\boldsymbol{v}} s.\end{aligned}$$

Pressure and temperature calculated from $\rho(n, s)$

Equation (22.6) lends itself to interpretation in two opposite senses: as a way to deduce the density of mass-energy of the medium from information about pressure (as a function of n and s) and temperature (as a function of n and s); and conversely, as a way to deduce the two functions $p(n, s)$ and $T(n, s)$ from the one function $\rho(n, s)$. It is natural to look at the second approach first; who does not like a strategy that makes an intellectual profit? Regarding ρ as a known (or calculable) function of n and s, one deduces from (22.6)

$$\frac{\rho + p}{n} = \left(\frac{\partial \rho}{\partial n}\right)_s,$$

$$nT = \left(\frac{\partial \rho}{\partial s}\right)_n,$$

and thence pressure and temperature individually,

$$p(n, s) = n\left(\frac{\partial \rho}{\partial n}\right)_s - \rho, \tag{22.7a}$$

$$T(n, s) = \frac{1}{n}\left(\frac{\partial \rho}{\partial s}\right)_n \tag{22.7b}$$

("two equations of state from one"). The analysis simplifies still further when the fluid, already assumed to be everywhere of the same composition, is also everywhere endowed with the same entropy per baryon, s, and is in a state of adiabatic flow (no shocks or heat conduction). Then the density $\rho = \rho(n, s)$ reduces to a function of one variable out of which one derives everything (ρ, p, μ) needed for the hydrodynamics and the gravitation physics of the system (next chapter). Other choices of the "primary thermodynamic potential" are appropriate under other circumstances (see Box 22.1).

If differentiation leads from $\rho(n, s)$ to $p(n, s)$ and $T(n, s)$, it does not follow that one can take any two functions $p(n, s)$ and $T(n, s)$ and proceed "backwards" (by integration) to the "primary function", $\rho(n, s)$. To be compatible with the first law of thermodynamics (22.6), the two functions must satisfy the consistency requirement ["Maxwell relation"; equality of second partial derivatives of ρ]

Maxwell relation

$$(\partial p/\partial s)_n = n^2(\partial T/\partial n)_s. \tag{22.7c}$$

Box 22.1 PRINCIPAL ALTERNATIVES FOR "PRIMARY THERMODYNAMIC POTENTIAL" TO DESCRIBE A FLUID

Primary thermodynamic potential and quantities on which it is most appropriately envisaged to depend	"Secondary" thermodynamic quantities obtained by differentiation of primary with or without use of $d\left(\frac{\rho}{n}\right) + p\,d\left(\frac{1}{n}\right) - T\,ds = 0$	Conditions under which convenient, appropriate, and relevant
"Density"; total amount of mass-energy (rest + thermal + $\cdots$) per unit volume $\rho = \rho(n, s)$	$p(n, s) = n\left(\frac{\partial\rho}{\partial n}\right)_s - \rho$ $T(n, s) = \frac{1}{n}\left(\frac{\partial\rho}{\partial s}\right)_n$ $\mu(n, s) = \frac{p+\rho}{n} = \left(\frac{\partial\rho}{\partial n}\right)_s$	Conditions of adiabatic flow (no shocks or heat conduction), so that s stays constant along streamline
"Physical free energy" $a(n, T) = \frac{\rho}{n} - Ts$	$p(n, T) = n^2\left(\frac{\partial a}{\partial n}\right)_T$ $s(n, T) = -\left(\frac{\partial a}{\partial T}\right)_n$ $\rho(n, T) = -nT^2\left[\frac{\partial(a/T)}{\partial T}\right]_n$	Know or can calculate a (or the "sum over states" of statistical mechanics) for conditions of specified volume per baryon and temperature
"Chemical free energy" $f(p, T) = \frac{\rho + p}{n} - Ts$	$1/n(p, T) = (\partial f/\partial p)_T$ $s(p, T) = -(\partial f/\partial T)_p$ $\rho(p, T) = \frac{f - T(\partial f/\partial T)_p}{(\partial f/\partial p)_T} - p$	Relevant for determining equilibrium when pressure and temperature are specified
"Chemical potential" ("energy to inject" expressed on a "per baryon" basis) $\mu(p, s) = \frac{p+\rho}{n}$	$1/n(p, s) = (\partial\mu/\partial p)_s$ $T(p, s) = (\partial\mu/\partial s)_p$ $\rho(p, s) = \frac{\mu}{(\partial\mu/\partial p)_s} - p$	When injection energy [= Fermi energy for an ideal Fermi gas, relativistic or not; see exercise 22.3] is the center of attention

2

Chemical potential equals "injection energy" at fixed entropy per baryon and total volume

The chemical potential μ is also a unique function of n and s. It is defined as follows. (1) Take a sample of the simple fluid in a fixed thermodynamic state (fixed n and s). (2) Take, separately, a much smaller sample of the same fluid, containing δA baryons in the same thermodynamic state as the large sample (same n and s). (3) Inject the smaller sample into the larger one, holding the volume of the large sample fixed during the injection process. (4) The total mass-energy injected,

$$\delta M_{\text{injected}} = \rho \times (\text{volume of injected fluid}) = \rho(\delta A/n),$$

plus the work required to perform the injection

$$\begin{aligned}\delta W_{\text{injection}} &= \begin{pmatrix}\text{work done against pressure of large sample}\\ \text{to open up space in it for the injected fluid}\end{pmatrix}\\ &= p(\text{volume of injected fluid}) = p(\delta A/n),\end{aligned}$$

is equal to $\mu\,\delta A$:

$$\mu\,\delta A = \delta M_{\text{injected}} + \delta W_{\text{injection}} = \frac{\rho + p}{n}\,\delta A.$$

Stated more briefly:

$$\begin{aligned}\mu &= \begin{pmatrix}\text{total mass-energy required, per baryon, to "create" and}\\ \text{inject a small additional amount of fluid into a given}\\ \text{sample, without changing } s \text{ or volume of the sample}\end{pmatrix}\\ &= \frac{\rho + p}{n} \underset{\text{[by first law of thermodynamics (22.6)]}}{=} \left(\frac{\partial \rho}{\partial n}\right)_s.\end{aligned} \tag{22.8}$$

All the above laws and equations of thermodynamics are the same in curved spacetime as in flat spacetime; and the same in (relativistic) flat spacetime as in classical nonrelativistic thermodynamics—except for the inclusion of rest mass, together with all other forms of mass-energy, in ρ and μ. The reason is simple: the laws are all formulated as scalar equations linking thermodynamic variables that one measures in the rest frame of the fluid.

§22.3. HYDRODYNAMICS IN CURVED SPACETIME*

Laws of hydrodynamics for simple fluid without heat flow or viscosity:

A simple perfect fluid flows through spacetime. It might be the Earth's atmosphere circulating in the Earth's gravitational field. It might be the gaseous interior of the Sun at rest in its own gravitational field. It might be interstellar gas accreting onto a black hole. But whatever and wherever the fluid may be, its motion will be governed by the curved-spacetime laws of thermodynamics (§22.2) plus the local

*For more detailed treatments of this subject see, e.g., Ehlers (1961), Taub (1971), Ellis (1971), Lichnerowicz (1967), Cattaneo (1971), and references cited therein; see also the references on kinetic theory cited at the beginning of §22.6.

law of energy-momentum conservation, $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$. The chief objective of this section is to reduce the equation $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ to usable form. The reduction will be performed in the text using abstract notation; the reader is encouraged to repeat the reduction using index notation.

(1) Origins of laws

The stress-energy tensor for a perfect fluid, in curved spacetime as in flat (equivalence principle!), is

$$\boldsymbol{T} = (\rho + p)\boldsymbol{u} \otimes \boldsymbol{u} + p\boldsymbol{g}. \tag{22.9}$$

(See §5.5.) Its divergence is readily calculated using the chain rule; using the "compatibility relation between $\boldsymbol{g}$ and $\boldsymbol{\nabla}$," $\boldsymbol{\nabla g} = 0$; using the identity $(\boldsymbol{\nabla} p) \cdot \boldsymbol{g} = \boldsymbol{\nabla} p$ (which one readily verifies in index notation); and using

$$\begin{aligned} 0 &= \underset{\uparrow \text{[divergence on first slot]}}{\boldsymbol{\nabla}} \cdot \boldsymbol{T} = [\boldsymbol{\nabla}(\rho + p) \cdot \boldsymbol{u}]\boldsymbol{u} + [(\rho + p)\boldsymbol{\nabla} \cdot \boldsymbol{u}]\boldsymbol{u} + [(\rho + p)\boldsymbol{u}] \cdot \boldsymbol{\nabla u} + (\boldsymbol{\nabla} p) \cdot \boldsymbol{g} \\ &= [\boldsymbol{\nabla}_{\boldsymbol{u}}\rho + \boldsymbol{\nabla}_{\boldsymbol{u}} p + (\rho + p)\boldsymbol{\nabla} \cdot \boldsymbol{u}]\boldsymbol{u} + (\rho + p)\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} + \boldsymbol{\nabla} p. \end{aligned} \tag{22.10}$$

The component of this equation along the 4-velocity is especially simple (recall that $\boldsymbol{u} \cdot \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = \frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u}^2 = 0$ because $\boldsymbol{u}^2 \equiv -1$):

$$\begin{aligned} 0 = \boldsymbol{u} \cdot (\boldsymbol{\nabla} \cdot \boldsymbol{T}) &= -[\boldsymbol{\nabla}_{\boldsymbol{u}}\rho + \boldsymbol{\nabla}_{\boldsymbol{u}} p + (\rho + p)\boldsymbol{\nabla} \cdot \boldsymbol{u}] + \boldsymbol{\nabla}_{\boldsymbol{u}} p \\ &= -\boldsymbol{\nabla}_{\boldsymbol{u}}\rho - (\rho + p)\boldsymbol{\nabla} \cdot \boldsymbol{u}. \end{aligned}$$

Combine this with the equation of baryon conservation (22.3); the result is

(2) Local energy conservation: adiabaticity of flow

$$\frac{d\rho}{d\tau} = \frac{(\rho + p)}{n}\frac{dn}{d\tau}. \tag{22.11a}$$

Notice that this is identical to the first law of thermodynamics (22.6) applied along a flow line, plus the assumption that the entropy per baryon is conserved along a flow line

$$ds/d\tau = 0. \tag{22.11b}$$

There is no reason for surprise at this result. To insist on thermodynamic equilibrium and to demand that the entropy remain constant is to require zero exchange of heat between one element of the fluid and another. But the stress-energy tensor (22.9) recognizes that heat exchange is absent. Any heat exchange would show up as an energy flux term in $\boldsymbol{T}$ (Ex. 22.7); but no such term is present. Consequently, when one studies local energy conservation by evaluating $\boldsymbol{u} \cdot (\boldsymbol{\nabla} \cdot \boldsymbol{T}) = 0$, the stress-energy tensor reports that no heat flow is occurring—i.e. that $ds/d\tau = 0$.

Three components of $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$ remain: the components orthogonal to the fluid's 4-velocity. One can pluck them out of $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$, leaving behind the component along $\boldsymbol{u}$, by use of the "projection tensor"

$$\boldsymbol{P} \equiv \boldsymbol{g} + \boldsymbol{u} \otimes \boldsymbol{u}. \tag{22.12}$$

2

Box 22.2 THERMODYNAMICS AND HYDRODYNAMICS FOR A SIMPLE PERFECT FLUID IN CURVED SPACETIME

A. Ten Quantities Characterize the Fluid

Thermodynamic potentials all measured in rest frame
- n, baryon number density
- ρ, density of total mass-energy
- p, pressure
- T, temperature
- s, entropy per baryon
- μ, chemical potential per baryon

Four components of the fluid 4-velocity

B. Ten Equations Govern the Fluid's Motion

Two *equations of state*

$$p = p(n, s), \qquad T = T(n, s) \qquad (1), (2)$$

subject to the compatibility constraint ("Maxwell relation," which follows from first law of thermodynamics)

$$(\partial p/\partial s)_n = n^2(\partial T/\partial n)_s.$$

First law of thermodynamics

$$\boldsymbol{d}\rho = \frac{\rho + p}{n}\,\boldsymbol{d}n + nT\,\boldsymbol{d}s, \qquad (3)$$

which can be integrated to give $\rho(n, s)$.

Equation for chemical potential

$$\mu = (\rho + p)/n, \qquad (4)$$

which can be combined with $\rho(n, s)$ and $p(n, s)$ to give $\mu(n, s)$.

Law of baryon conservation

$$dn/d\tau \equiv \boldsymbol{\nabla}_{\boldsymbol{u}} n = -n\,\boldsymbol{\nabla}\cdot\boldsymbol{u}. \qquad (5)$$

Conservation of energy along flow lines, which (assuming no energy exchange between adjacent fluid elements) means "*adiabatic flow*"

$$ds/d\tau = 0 \text{ except in shock waves, where } ds/d\tau > 0. \qquad (6)$$

[Shock waves are not treated in this book; see Taub (1948), de Hoffman and Teller (1950), Israel (1960), May and White (1967), Zel'dovich and Rayzer (1967); Lichnerowicz (1967, 1971); and Thorne (1973a).]

Euler equations

$$(\rho + p)\,\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = -(\boldsymbol{g} + \boldsymbol{u}\otimes\boldsymbol{u})\cdot\boldsymbol{\nabla} p, \qquad (7), (8), (9)$$

which determine the flow lines to which $\boldsymbol{u}$ is tangent.

Normalization of 4-velocity

$$\boldsymbol{u}\cdot\boldsymbol{u} = -1. \qquad (10)$$

(See exercise 22.4.) Contracting $\boldsymbol{P}$ with $\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$ [equation (22.10)] gives

(3) Euler equation

$$(\rho + p)\,\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = -\boldsymbol{P}\cdot(\boldsymbol{\nabla} p) \equiv -[\boldsymbol{\nabla} p + (\boldsymbol{\nabla}_{\boldsymbol{u}} p)\boldsymbol{u}]. \qquad (22.13)$$

This is the *"Euler equation"* of relativistic hydrodynamics. It has precisely the same form as the corresponding flat-spacetime Euler equation:

$$\begin{pmatrix}\text{inertial mass}\\ \text{per unit volume}\\ \text{[exercise 5.4]}\end{pmatrix} \times \begin{pmatrix}\text{4-acceleration}\\ \text{of fluid}\end{pmatrix} = -\begin{pmatrix}\text{pressure gradient}\\ \text{in the 3-surface}\\ \text{orthogonal to 4-velocity}\end{pmatrix}. \qquad (22.13')$$

The pressure gradient, not "gravity," is responsible for all deviation of flow lines from geodesics.

Box 22.2 reorganizes and summarizes the above laws of thermodynamics and hydrodynamics.

EXERCISES

Exercise 22.1. DIVERGENCE OF FLOW LINES PRODUCES VOLUME CHANGES

Derive the equation $dV/d\tau = (\nabla \cdot \boldsymbol{u})V$ [equation (22.2)] for the rate of change of volume of a fluid element. [*Hint:* Pick an event $\mathscr{P}_0$, and calculate in a local Lorentz frame at $\mathscr{P}_0$ which momentarily moves with the fluid ("rest frame at $\mathscr{P}_0$").] [*Solution:* At events near $\mathscr{P}_0$ the fluid has a very small ordinary velocity $v^j = dx^j/dt$. Consequently a cube of fluid at $\mathscr{P}_0$ with edges $\Delta x = \Delta y = \Delta z = L$ changes its edges, after time δt, by the amounts

$$\begin{aligned}\delta(\Delta x) &= [(dx/dt)\,\delta t]_{\text{at "front face"}} - [(dx/dt)\,\delta t]_{\text{at "back face"}}\\ &= (\partial v^x/\partial x)L\,\delta t,\\ \delta(\Delta y) &= (\partial v^y/\partial y)L\,\delta t,\\ \delta(\Delta z) &= (\partial v^z/\partial z)L\,\delta t.\end{aligned}$$

The corresponding change in volume is

$$\delta(\Delta x\,\Delta y\,\Delta z) = (\partial v^j/\partial x^j)L^3\,\delta t;$$

so the rate of change of volume is

$$\partial V/\partial t = V(\partial v^j/\partial x^j).$$

But in the local Lorentz rest frame at and near $\mathscr{P}_0$ (where $x^\alpha = 0$), the metric coefficients are $g_{\mu\nu} = \eta_{\mu\nu} + 0(|x^\alpha|^2)$, and the ordinary velocity is $v^j = 0(|x^\alpha|)$; so

$$u^0 = \frac{dt}{d\tau} = \frac{dt}{(-g_{\mu\nu}\,dx^\mu\,dx^\nu)^{1/2}} = 1 + 0(|x^\alpha|^2),$$

$$u^j = \frac{dx^j}{d\tau} = v^j + 0(|x^\alpha|^3).$$

Thus, the derivatives $\partial V/\partial t$ and $V(\partial v^j/\partial x^j)$ at $\mathscr{P}_0$ are

$$\begin{aligned}\partial V/\partial t &= u^\alpha \partial V/\partial x^\alpha = u^\alpha V_{,\alpha} = dV/d\tau\\ &= V(\partial v^j/\partial x^j) = V(\partial u^\alpha/\partial x^\alpha) = Vu^\alpha{}_{;\alpha} = V(\nabla \cdot \boldsymbol{u}). \qquad \text{Q.E.D.]}\end{aligned}$$

[Note that by working in flat spacetime, one could have inferred more easily that $\partial V/\partial t = dV/d\tau$ and $\partial v^j/\partial x^j = \nabla \cdot \boldsymbol{u}$; one would then have concluded $dV/d\tau = (\nabla \cdot \boldsymbol{u})V$; and one could have invoked the equivalence principle to move this law into curved spacetime.]

Exercise 22.2. EQUATION OF CONTINUITY

Show that in the nonrelativistic limit in flat spacetime the equation of baryon conservation (22.3) becomes the "equation of continuity"

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x^j}(nv^j) = 0.$$

Exercise 22.3. CHEMICAL POTENTIAL FOR IDEAL FERMI GAS

Show that the chemical potential of an ideal Fermi gas, nonrelativistic or relativistic, is (at zero temperature) equal to the Fermi energy (energy of highest occupied momentum state) of that gas.

Exercise 22.4. PROJECTION TENSORS

Show that contraction of a tangent vector $\boldsymbol{B}$ with the "projection tensor" $\boldsymbol{P} \equiv \boldsymbol{g} + \boldsymbol{u} \otimes \boldsymbol{u}$ projects $\boldsymbol{B}$ into the 3-surface orthogonal to the 4-velocity vector $\boldsymbol{u}$. [Hint: perform the calculation in an orthonormal frame with $\boldsymbol{e}_{\hat{0}} = \boldsymbol{u}$, and write $\boldsymbol{B} = B^{\alpha}\boldsymbol{e}_{\hat{\alpha}}$; then show that $\boldsymbol{P} \cdot \boldsymbol{B} = B^{\hat{j}}\boldsymbol{e}_{\hat{j}}$.] If $\boldsymbol{n}$ is a unit spacelike vector, show that $\boldsymbol{P} \equiv \boldsymbol{g} - \boldsymbol{n} \otimes \boldsymbol{n}$ is the corresponding projection operator. Note: There is no *unique* concept of "the projection orthogonal to a null vector." Why? [Hint: draw pictures in flat spacetime suppressing one spatial dimension.]

Exercise 22.5. PRESSURE GRADIENT IN STATIONARY GRAVITATIONAL FIELD

A perfect fluid is at rest (flow lines have $x^j =$ constant) in a stationary gravitational field (metric coefficients are independent of x^0). Show that the pressure gradient required to "support the fluid against gravity" (i.e., to make its flow lines be $x^j =$ constant instead of geodesics) is

$$\frac{\partial p}{\partial x^0} = 0, \qquad \frac{\partial p}{\partial x^j} = -(\rho + p)\frac{\partial \ln \sqrt{-g_{00}}}{\partial x^j}. \tag{22.14}$$

Evaluate this pressure gradient in the Newtonian limit, using the coordinate system and metric coefficients of equation (18.15c).

Exercise 22.6. EXPANSION, ROTATION, AND SHEAR

Let a field of fluid 4-velocities $\boldsymbol{u}(\mathscr{P})$ be given.

(a) Show that $\boldsymbol{\nabla u}$ can be decomposed in the following manner:

$$u_{\alpha;\beta} = \omega_{\alpha\beta} + \sigma_{\alpha\beta} + \frac{1}{3}\theta P_{\alpha\beta} - a_\alpha u_\beta, \tag{22.15a}$$

where $\boldsymbol{a}$ is the *4-acceleration* of the fluid

$$a_\alpha \equiv u_{\alpha;\beta}u^\beta, \tag{22.15b}$$

θ is the *"expansion"* of the fluid world lines

$$\theta \equiv \boldsymbol{\nabla} \cdot \boldsymbol{u} = u^\alpha{}_{;\alpha}, \tag{22.15c}$$

$P_{\alpha\beta}$ is the *projection tensor*

$$P_{\alpha\beta} \equiv g_{\alpha\beta} + u_\alpha u_\beta, \tag{22.15d}$$

$\sigma_{\alpha\beta}$ is the *shear tensor* of the fluid

$$\sigma_{\alpha\beta} \equiv \frac{1}{2}(u_{\alpha;\mu}P^\mu{}_\beta + u_{\beta;\mu}P^\mu{}_\alpha) - \frac{1}{3}\theta P_{\alpha\beta}, \tag{22.15e}$$

and $\omega_{\alpha\beta}$ is the *rotation 2-form* of the fluid

$$\omega_{\alpha\beta} \equiv \frac{1}{2}(u_{\alpha;\mu}P^\mu{}_\beta - u_{\beta;\mu}P^\mu{}_\alpha). \tag{22.15f}$$

(b) Each of the component parts of this decomposition has a simple physical interpretation in the local rest frames of the fluid. The interpretation of the 4-acceleration $\boldsymbol{a}$ in terms of accelerometer readings should be familiar. Exercise 22.1 showed that the expansion $\theta = \boldsymbol{\nabla} \cdot \boldsymbol{u}$ describes the rate of increase of the volume of a fluid element,

$$\theta = (1/V)(dV/d\tau). \tag{22.15g}$$

Exercise 22.4 explored the meaning and use of the projection tensor $\boldsymbol{P}$. Verify that in a local Lorentz frame ($g_{\hat{\alpha}\hat{\beta}} = \eta_{\alpha\beta}$, $\Gamma^{\hat{\alpha}}{}_{\hat{\beta}\hat{\gamma}} = 0$) momentarily moving with the fluid ($u^{\hat{\alpha}} = \delta^\alpha{}_0$), $\sigma_{\hat{\alpha}\hat{\beta}}$ and $\omega_{\hat{\alpha}\hat{\beta}}$ reduce to the classical (nonrelativistic) shear and rotation of the fluid. [See, e.g., §§2.4 and 2.5 of Ellis (1971) for both classical and relativistic descriptions of shear and rotation.]

Exercise 22.7. HYDRODYNAMICS WITH VISCOSITY AND HEAT FLOW.*

(a) In §15 of Landau and Lifshitz (1959), one finds an analysis of viscous stresses for a classical (nonrelativistic) fluid. By carrying that analysis over directly to the local Lorentz rest frame of a relativistic fluid, and by then generalizing to frame-independent language, show that the contribution of viscosity to the stress-energy tensor is

$$\boldsymbol{T}^{(\text{visc})} = -2\eta\boldsymbol{\sigma} - \zeta\theta\boldsymbol{P}, \tag{22.16a}$$

where $\eta \geq 0$ is the *"coefficient of dynamic viscosity"*; $\zeta \geq 0$ is the *"coefficient of bulk viscosity"*; and $\boldsymbol{\sigma}$, θ, $\boldsymbol{P}$ are the shear, expansion, and projection tensor of the fluid.

(b) An idealized description of heat flow in a fluid introduces the *heat-flux 4-vector* $\boldsymbol{q}$ with components in the local rest-frame of the fluid,

$$q^{\hat{0}} = 0, \qquad q^{\hat{j}} = \begin{pmatrix}\text{energy per unit time crossing unit}\\ \text{surface perpendicular to } \boldsymbol{e}_{\hat{j}}\end{pmatrix}. \tag{22.16b}$$

By generalizing from the fluid rest frame to frame-independent language, show that the contribution of heat flux to the stress-energy tensor is

$$\boldsymbol{T}^{(\text{heat})} = \boldsymbol{u} \otimes \boldsymbol{q} + \boldsymbol{q} \otimes \boldsymbol{u}. \tag{22.16c}$$

Thereby conclude that, in this idealized picture, the stress-energy tensor for a viscous fluid with heat conduction is

$$T^{\alpha\beta} = \rho u^\alpha u^\beta + (p - \zeta\theta)P^{\alpha\beta} - 2\eta\sigma^{\alpha\beta} + q^\alpha u^\beta + u^\alpha q^\beta. \tag{22.16d}$$

(c) Define the entropy 4-vector $\boldsymbol{s}$ by

$$\boldsymbol{s} \equiv ns\boldsymbol{u} + \boldsymbol{q}/T. \tag{22.16e}$$

By calculations in the local rest-frame of the fluid, show that

$$\begin{aligned}\boldsymbol{\nabla}\cdot\boldsymbol{s} &= \begin{pmatrix}\text{rate of increase of entropy}\\ \text{in a unit volume}\end{pmatrix} - \begin{pmatrix}\text{rate at which heat and fluid}\\ \text{carry entropy into a unit volume}\end{pmatrix}\\ &= \begin{pmatrix}\text{rate at which entropy is being}\\ \text{generated in a unit volume}\end{pmatrix}.\end{aligned} \tag{22.16f}$$

Thereby arrive at the following form of the *second law of thermodynamics:*

$$\boldsymbol{\nabla}\cdot\boldsymbol{s} \geq 0. \tag{22.16g}$$

(d) Calculate the law of local energy conservation, $\boldsymbol{u}\cdot\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$, for a viscous fluid with heat flow. Combine with the first law of thermodynamics and with the law of baryon conservation to obtain

$$T\boldsymbol{\nabla}\cdot\boldsymbol{s} = \zeta\theta^2 + 2\eta\sigma_{\alpha\beta}\sigma^{\alpha\beta} - q^\alpha(T_{,\alpha}/T + a_\alpha). \tag{22.16h}$$

Interpret each term of this equation as a contribution to entropy generation (*example:* $2\eta\sigma_{\alpha\beta}\sigma^{\alpha\beta}$ describes entropy generation by viscous heating). [*Note:* The term $q^\alpha a_\alpha$ is relativistic in origin. It is associated with the inertia of the flowing heat.]

(e) When one takes account of the inertia of the flowing heat, one obtains the following generalization of the classical law of heat conduction:

$$q^\alpha = -\kappa P^{\alpha\beta}(T_{,\beta} + Ta_\beta) \tag{22.16i}$$

*Exercise supplied by John M. Stewart.

(Eckart 1940). Here κ is the *coefficient of thermal conductivity*. Use this equation to show that, for a fluid at rest in a stationary gravitational field (Exercise 22.5),

$$q_0 = 0, \qquad q_j = -\frac{\kappa}{\sqrt{-g_{00}}}(T\sqrt{-g_{00}})_{,j}. \tag{22.16j}$$

[Thus, thermal equilibrium corresponds not to constant temperature, but to the redshifted temperature distribution $T\sqrt{-g_{00}}$ = constant; Tolman (1934a), p. 313.] Also, use the idealized law of heat conduction (22.16i) to reexpress the rate of entropy generation as

$$\begin{aligned} T\boldsymbol{\nabla}\cdot\boldsymbol{s} &= \zeta\theta^2 + 2\eta\sigma_{\alpha\beta}\sigma^{\alpha\beta} + (\kappa/T)P^{\alpha\beta}(T_{,\alpha} + Ta_\alpha)(T_{,\beta} + Ta_\beta) \\ &\geq 0. \end{aligned} \tag{22.16k}$$

[For further details about heat flow and for discussions of the limitations of the above idealized description, see e.g., §4.18 of Ehlers (1971); also Marle (1969), Anderson (1970), Stewart (1971), and papers cited therein.]

---

§22.4. ELECTRODYNAMICS IN CURVED SPACETIME

Electric and magnetic fields

In a local Lorentz frame in the presence of gravity, an observer can measure the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ using the usual Lorentz force law for charged particles. As in special relativity, he can regard $\boldsymbol{E}$ and $\boldsymbol{B}$ as components of an electromagnetic field tensor,

$$F^{\hat{0}\hat{j}} = -F^{\hat{j}\hat{0}} = E^{\hat{j}}, \qquad F^{\hat{j}\hat{k}} = \epsilon^{\hat{j}\hat{k}\hat{l}}B^{\hat{l}};$$

he can regard the charge and current densities as components of a 4-vector $J^{\hat{\alpha}}$, and he can write Maxwell's equations and the Lorentz force equation in the special relativistic form,

$$F^{\hat{\alpha}\hat{\beta}}{}_{,\hat{\beta}} = 4\pi J^{\hat{\alpha}}, \qquad F_{\hat{\alpha}\hat{\beta},\hat{\gamma}} + F_{\hat{\beta}\hat{\gamma},\hat{\alpha}} + F_{\hat{\gamma}\hat{\alpha},\hat{\beta}} = 0,$$

$$ma^{\hat{\alpha}} = F^{\hat{\alpha}\hat{\beta}}qu_{\hat{\beta}} \qquad \begin{pmatrix} m = \text{mass of particle, } q = \text{charge,} \\ u^{\hat{\alpha}} = \text{4-velocity, } a^{\hat{\alpha}} = \text{4-acceleration} \end{pmatrix}.$$

In any other frame these equations will have the same form, but with commas replaced by semicolons

Maxwell equations and Lorentz force law

$$F^{\alpha\beta}{}_{;\beta} = 4\pi J^\alpha, \tag{22.17a}$$

$$F_{\alpha\beta;\gamma} + F_{\beta\gamma;\alpha} + F_{\gamma\alpha;\beta} = 0, \tag{22.17b}$$

$$ma^\alpha = F^{\alpha\beta}qu_\beta. \tag{22.17c}$$

These are the basic equations of electrodynamics in the presence of gravity. From them follows everything else. For example, as in special relativity, so also here (exercise 22.9), they imply the equation of charge conservation

Charge conservation

$$J^\alpha{}_{;\alpha} = 0; \tag{22.18a}$$

and for an electromagnetic field interacting with charged matter (exercise 22.10) they imply vanishing divergence for the sum of the stress-energy tensors

Local conservation of energy-momentum

$$(T^{(\mathrm{EM})\alpha\beta} + T^{(\mathrm{MATTER})\alpha\beta})_{;\beta} = 0. \tag{22.18b}$$

Vector potential

As in special relativity, so also here, one can introduce a vector potential A^μ. Replacing commas by semicolons in the usual special-relativistic expression for $F^{\mu\nu}$ in terms of A^μ, one obtains

$$F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu}. \tag{22.19a}$$

If all is well, this equation should guarantee (as in special relativity) that the Maxwell equations (22.17b) are satisfied. Indeed, it does, as one sees in exercise 22.8. To derive the wave equation that governs the vector potential, insert expression (22.19a) into the remaining Maxwell equations (22.17a), obtaining

$$-A^{\alpha;\beta}{}_{\beta} + A^{\beta;\alpha}{}_{\beta} = 4\pi J^\alpha; \tag{22.19b}$$

then commute covariant derivatives in the first term using the identity (16.6c), to obtain

$$-A^{\alpha;\mu}{}_{\mu} + A^{\mu}{}_{;\mu}{}^{;\alpha} + R^\alpha{}_\mu A^\mu = 4\pi J^\alpha. \tag{22.19b$'$}$$

Finally, adopting the standard approach of special relativity, impose the Lorentz gauge condition

Lorentz gauge condition

$$A^\mu{}_{;\mu} = 0, \tag{22.19c}$$

thereby bringing the wave equation (22.19b′) into the form

Wave equation for vector potential

$$(\Delta_{dR} A)^\alpha \equiv - A^{\alpha;\beta}{}_{\beta} + R^\alpha{}_\beta A^\beta = 4\pi J^\alpha. \tag{22.19d}$$

The "de Rham vector wave operator" Δ which appears here is, apart from sign, a generalized d'Alambertian for vectors in curved spacetime. Mathematically it is more powerful than $-A^{\alpha;\beta}{}_{;\beta}$, and than any other operator that reduces to (minus) the d'Alambertian in special relativity. [For a discussion, see de Rham (1955).]

Although the electrodynamic equations (22.17a)–(22.19b) are all obtained from special relativity by the comma-goes-to-semicolon rule, the wave equation (22.19d) for the vector potential is not ("curvature coupling"; see Box 16.1). Nevertheless, when spacetime is flat (so $R^\alpha{}_\beta = 0$), (22.19d) does reduce to the usual wave equation of special relativity.

EXERCISES

Exercise 22.8. THE VECTOR POTENTIAL FOR ELECTRODYNAMICS

Show that in any coordinate frame the connection coefficients cancel out of both equations (22.19a) and (22.17b), so they can be written

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}, \tag{22.20a}$$

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0. \tag{22.20b}$$

(In the language of differential forms these equations are $\boldsymbol{F} = \boldsymbol{dA}$, $\boldsymbol{dF} = 0$.) Then use this form of the equations to show that equation (22.19a) implies equation (22.17b), as asserted in the text.

Exercise 22.9. CHARGE CONSERVATION IN THE PRESENCE OF GRAVITY

Show that Maxwell's equations (22.17a,b) imply the equation of charge conservation (22.18a) when gravity is present, just as they do in special relativity theory. [Hints: Use the antisymmetry of $F^{\alpha\beta}$; and beware of the noncommutation of the covariant derivatives, which must be handled using equations (16.6). Alternatively, show that in coordinate frames, equation (22.17a) can be written as

$$\frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\beta} (\sqrt{|g|} F^{\alpha\beta}) = 4\pi J^\alpha \tag{22.17a$'$}$$

and (22.18a) as

$$J^\alpha{}_{;\alpha} \equiv \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\alpha} (\sqrt{|g|} J^\alpha) = 0, \tag{22.18a$'$}$$

and carry out the demonstration in a coordinate frame.]

Exercise 22.10. INTERACTING ELECTROMAGNETIC FIELD AND CHARGED MATTER

As in special relativity, so also in the presence of gravity ("equivalence principle"), the stress-energy tensor for an electromagnetic field is

$$T^{(\mathrm{EM})}{}_{\alpha\beta} = \frac{1}{4\pi} \left(F_{\alpha\mu} F_\beta{}^\mu - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} g_{\alpha\beta} \right). \tag{22.21}$$

Use Maxwell's equations (22.17a,b) in the presence of gravity to show that

$$T^{(\mathrm{EM})\alpha\beta}{}_{;\beta} = -F^{\alpha\beta} J_\beta. \tag{22.22}$$

But $F^{\alpha\beta} J_\beta$ is just the Lorentz 4-force per unit volume with which the electromagnetic field acts on the charged matter [see the Lorentz force equation (22.17c); also equation (5.43)]; i.e., it is $T^{(\mathrm{MATTER})\alpha\beta}{}_{;\beta}$. Consequently, the above equation can be rewritten in the form (22.18b) cited in the text.

§22.5. GEOMETRIC OPTICS IN CURVED SPACETIME*

Radio waves from the quasar 3C279 pass near the sun and get deflected by its gravitational field. Light rays emitted by newborn galaxies long ago and far away propagate through the cosmologically curved spacetime of the universe, and get focused (and redshifted) producing curvature-enlarged (but dim) images of the galaxies on the Earth's sky.

*Based in part on notes prepared by William L. Burke at Caltech in 1968. For more detailed treatments of geometric optics in curved spacetime, see, e.g., Sachs (1961), Jordan, Ehlers, and Sachs (1961), and Robinson (1961); also references discussed and listed in §41.11.

These and most other instances of the propagation of light and radio waves are subject to the laws of geometric optics. This section derives those laws, in curved spacetime, from Maxwell's equations.

Overview of geometric optics

The fundamental laws of geometric optics are: (1) light rays are null geodesics; (2) the polarization vector is perpendicular to the rays and is parallel-propagated along the rays; and (3) the amplitude is governed by an adiabatic invariant which, in quantum language, states that the number of photons is conserved.

Conditions for validity of geometric optics

The conditions under which these laws hold are defined by conditions on three lengths: (1) the typical reduced wavelength of the waves,

$$\lambda\!\!\!^{-} \equiv \frac{\lambda}{2\pi} = \begin{pmatrix}\text{"classical distance of closest approach for} \\ \text{a photon with one unit of angular momentum"}\end{pmatrix}, \qquad (22.23\text{a})$$

as measured in a typical local Lorentz frame (e.g., a frame at rest relative to nearby galaxies); (2) the typical length $\mathcal{L}$ over which the amplitude, polarization, and wavelength of the waves vary, e.g., the radius of curvature of a wave front, or the length of a wave packet produced by a sudden outburst in a quasar; (3) the typical radius of curvature $\mathcal{R}$ of the spacetime through which the waves propagate,

$$\mathcal{R} \equiv \begin{vmatrix}\text{typical component of } \textbf{\textit{Riemann}} \text{ as measured} \\ \text{in typical local Lorentz frame}\end{vmatrix}^{-1/2}. \qquad (22.23\text{b})$$

Geometric optics is valid whenever the reduced wavelength is very short compared to each of the other scales present,

$$\lambda\!\!\!^{-} \ll \mathcal{L} \quad \text{and} \quad \lambda\!\!\!^{-} \ll \mathcal{R}, \qquad (22.23\text{c})$$

so that the waves can be regarded *locally* as plane waves propagating through spacetime of negligible curvature.

The "two-length-scale" expansion underlying geometric optics

Mathematically one exploits the geometric-optics assumption, $\lambda\!\!\!^{-} \ll \mathcal{L}$ and $\lambda\!\!\!^{-} \ll \mathcal{R}$, as follows. Focus attention on waves that are highly monochromatic over regions $\lesssim \mathcal{L}$. (More complex spectra can be analyzed by superposition, i.e., by Fourier analysis.) Split the vector potential of electromagnetic theory into a rapidly changing, real phase,

$$\theta \sim (\text{distance propagated})/\lambda\!\!\!^{-},$$

and a slowly changing, complex amplitude (i.e. one with real and imaginary parts),

$$\boldsymbol{A} = \text{Real part of}\{\text{amplitude} \times e^{i\theta}\} \equiv \mathfrak{R}\{\text{amplitude} \times e^{i\theta}\}.$$

Imagine holding fixed the scale of the amplitude variation, $\mathcal{L}$, and the scale of the spacetime curvature, $\mathcal{R}$, while making the reduced wavelength, $\lambda\!\!\!^{-}$, shorter and shorter. The phase will get larger and larger ($\theta \propto 1/\lambda\!\!\!^{-}$) at any fixed event in spacetime, but the amplitude as a function of location in spacetime can remain virtually unchanged,

$$\text{Amplitude} = \begin{bmatrix}\text{dominant part,} \\ \text{independent of } \lambda\!\!\!^{-}\end{bmatrix} + \begin{bmatrix}\text{small corrections (deviations from} \\ \text{geometric optics) due to finite wavelength}\end{bmatrix}.$$

This circumstance allows one to expand the amplitude in powers of $\bar{\lambda}$:*

$$\text{Amplitude} = \underbrace{\boldsymbol{a}}_{\left[\text{independent of } \bar{\lambda}\right]} + \underbrace{\boldsymbol{b}}_{[\propto \bar{\lambda}]} + \underbrace{\boldsymbol{c}}_{[\propto \bar{\lambda}^2]} + \ldots .$$

[Actually, the expansion proceeds in powers of the dimensionless number

$$\bar{\lambda}/(\text{minimum of } \mathcal{L} \text{ and } \mathcal{R}) \equiv \bar{\lambda}/L. \tag{22.24}$$

Applied mathematicians call this a "two-length-scale expansion"; see, e.g., Cole (1968). The basic short-wavelength approximation here has a long history; see, e.g., Liouville (1837), Rayleigh (1912). Following a suggestion of Debye, it was applied to Maxwell's equations by Sommerfeld and Runge (1911). It is familiar as the WKB approximation in quantum mechanics, and has many other applications as indicated by the bibliography in Keller, Lewis, and Seckler (1956). The contribution of higher order terms is considered by Kline (1954) and Lewis (1958). See especially the book of Fröman and Fröman (1965).]

It is useful to introduce a parameter ε that keeps track of how rapidly various terms approach zero (or infinity) as $\bar{\lambda}/L$ approaches zero:

The vector potential in geometric optics

$$A_\mu = \Re\{(a_\mu + \varepsilon b_\mu + \varepsilon^2 c_\mu + \cdots)e^{i\theta/\varepsilon}\}. \tag{22.25}$$

Any term with a factor ε^n in front of it varies as $(\bar{\lambda}/L)^n$ in the limit of very small wavelengths [$\theta \propto (\bar{\lambda}/L)^{-1}$; $c_\mu \propto (\bar{\lambda}/L)^2$; etc.]. By convention, ε is a dummy expansion parameter with eventual value unity; so it can be dropped from the calculations when it ceases to be useful. And by convention, all "post-geometric-optics corrections" are put into the amplitude terms $\boldsymbol{b}, \boldsymbol{c}, \ldots$; none are put into θ.

Note that, while the phase θ is a real function of position in spacetime, the amplitude and hence the vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots$ are complex. For example, to describe monochromatic waves with righthand circular polarization, propagating in the z direction, one could set $\theta = \omega(z - t)$ and $\boldsymbol{a} = 1/\sqrt{2}a(\boldsymbol{e}_x + i\boldsymbol{e}_y)$ with a real; so

$$\boldsymbol{A} = \Re\left\{\frac{1}{\sqrt{2}}a(\boldsymbol{e}_x + i\boldsymbol{e}_y)e^{i\omega(z-t)}\right\} = \frac{1}{\sqrt{2}}a\{\cos[\omega(z-t)]\boldsymbol{e}_x - \sin[\omega(z-t)]\boldsymbol{e}_y\}$$

The assumed form (22.25) for the vector potential is the mathematical foundation of geometric optics. All the key equations of geometric optics result from inserting this vector potential into the source-free wave equation $\Delta\boldsymbol{A} = 0$ [equation (22.19d)] and into the Lorentz gauge condition $\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0$ [equation (22.19c)]. The resulting equations (derived below) take their simplest form only when expressed in terms of the following:

Basic concepts of geometric optics:

*The equations for $\boldsymbol{A}$ are linear. Therefore the analysis would proceed equally well assuming, instead of an amplitude independent of $\bar{\lambda}$, a dominant term $\boldsymbol{a} \propto \bar{\lambda}^n$, with $\boldsymbol{b} \propto \bar{\lambda}^{n+1}$, $\boldsymbol{c} \propto \bar{\lambda}^{n+2}$, etc. The results are independent of n. Choosing $n = 1$ would give field strengths $F_{\mu\nu}$ and energy densities $T_{\mu\nu} \propto F^2 \propto A^2/\bar{\lambda}^2 \propto$ constant as $\bar{\lambda} \longrightarrow 0$.

(1) wave vector

(2) scalar amplitude

(3) polarization vector

$$\text{“wave vector,”}\ \boldsymbol{k} \equiv \boldsymbol{\nabla}\theta; \tag{22.26a}$$

$$\text{“scalar amplitude,”}\ a \equiv (\boldsymbol{a} \cdot \bar{\boldsymbol{a}})^{1/2} = (a^\mu \bar{a}_\mu)^{1/2}; \tag{22.26b}$$

$$\text{“polarization vector,”}\ \boldsymbol{f} \equiv \boldsymbol{a}/a = \text{“unit complex vector along } \boldsymbol{a}\text{”}. \tag{22.26c}$$

(4) light rays

(Here $\bar{\boldsymbol{a}}$ is the complex conjugate of $\boldsymbol{a}$.) *Light rays* are defined to be the curves $\mathscr{P}(\lambda)$ normal to surfaces of constant phase θ. Since $\boldsymbol{k} \equiv \boldsymbol{\nabla}\theta$ is the normal to these surfaces, the differential equation for a light ray is

$$\frac{dx^\mu}{d\lambda} = k^\mu(x) = g^{\mu\nu}(x)\theta_{,\nu}(x). \tag{22.26d}$$

Box 22.3, appropriate for study at this point, shows the polarization vector, wave vector, surfaces of constant phase, and light rays for a propagating wave; the scalar amplitude, not shown there, merely tells the length of the vector amplitude $\boldsymbol{a}$. Insight into the complex polarization vector, if not familiar from electrodynamics, can be developed later in Exercise 22.12.

Derivation of laws of geometric optics

So much for the foundations. Now for the calculations. First insert the geometric-optics vector potential (22.25) into the Lorentz gauge condition:

$$0 = A^\mu{}_{;\mu} = \Re\left\{\left[\frac{i}{\varepsilon} k_\mu(a^\mu + \varepsilon b^\mu + \cdots) + (a^\mu + \varepsilon b^\mu + \cdots)_{;\mu}\right] e^{i\theta/\varepsilon}\right\}. \tag{22.27}$$

The leading term (order $1/\varepsilon$) says

$$\boldsymbol{k} \cdot \boldsymbol{a} = 0\ (\textit{amplitude is perpendicular to wave vector}); \tag{22.28}$$

or, equivalently

$$\boldsymbol{k} \cdot \boldsymbol{f} = 0\ (\textit{polarization is perpendicular to wave vector}). \tag{22.28$'$}$$

The post-geometric-optics breakdown in this orthogonality condition is governed by the higher-order terms $[0(1), 0(\varepsilon), 0(\varepsilon^2), \ldots]$ in the gauge condition (22.27); for example, the 0(1) terms say

$$\boldsymbol{k} \cdot \boldsymbol{b} = i\boldsymbol{\nabla} \cdot \boldsymbol{a}.$$

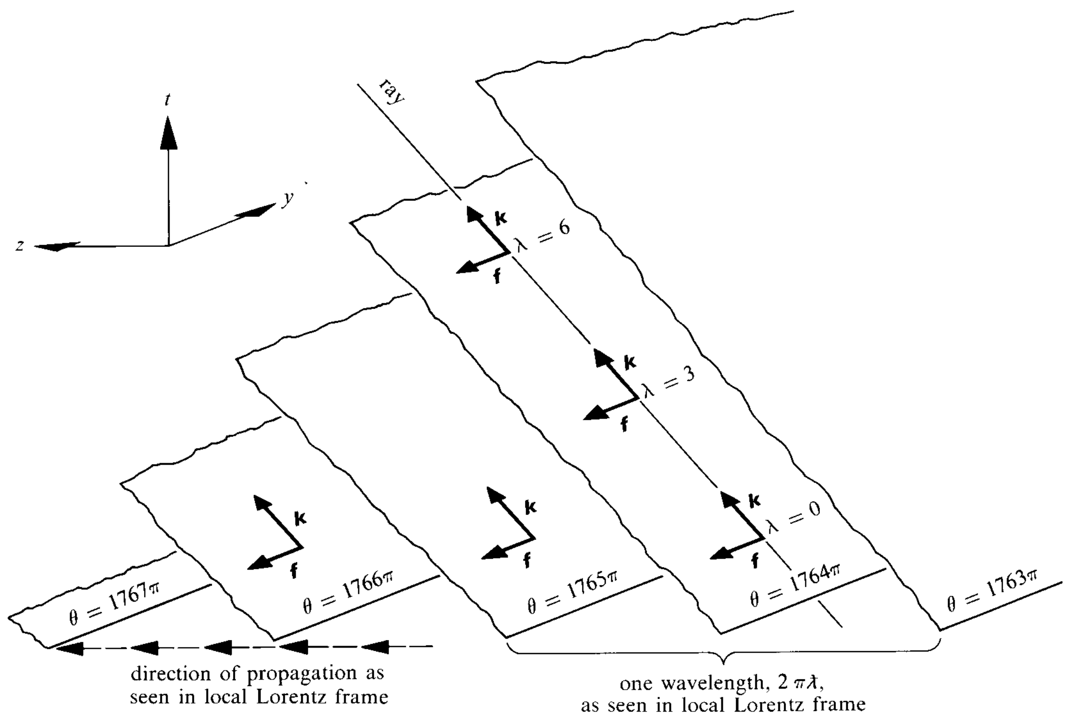
Next insert the vector potential (22.25) into the source-free wave equation (22.19d):

$$\begin{aligned} 0 = (\Delta_{dR}\boldsymbol{A})^\alpha &= -A^{\alpha;\beta}{}_\beta + R^\alpha{}_\beta A^\beta \\ &= \Re\left\{\left[\frac{1}{\varepsilon^2} k^\beta k_\beta(a^\alpha + \varepsilon b^\alpha + \varepsilon^2 c^\alpha + \cdots) - 2\frac{i}{\varepsilon} k^\beta(a^\alpha + \varepsilon b^\alpha + \cdots)_{;\beta}\right.\right. \\ &\quad \left.\left. - \frac{i}{\varepsilon} k^\beta{}_{;\beta}(a^\alpha + \varepsilon b^\alpha + \cdots) - (a^\alpha + \cdots)^{;\beta}{}_\beta + R^\alpha{}_\beta(a^\beta + \cdots)\right] e^{i\theta/\varepsilon}\right\}. \end{aligned} \tag{22.29}$$

Collect terms of order $1/\varepsilon^2$ and $1/\varepsilon$ (terms of order higher than $1/\varepsilon$ govern post-geometric-optics corrections):

(*continued on page 576*)

Box 22.3 GEOMETRY OF AN ELECTROMAGNETIC WAVE TRAIN



The drawing shows surfaces of constant phase, $\theta =$ constant, emerging through the "surface of simultaneity", $t = 0$, of a local Lorentz frame. The surfaces shown are alternately "crests" ($\theta = 1764\pi$, $\theta = 1766\pi$, . . .) and "troughs" ($\theta = 1765\pi$, $\theta = 1767\pi, \ldots$) of the wave train. These surfaces make up a 1-form, $\tilde{\boldsymbol{k}} = \boldsymbol{d}\theta$. The "corresponding vector" $\boldsymbol{k} = \boldsymbol{\nabla}\theta$ is the "*wave vector.*" The wave vector is null, $\boldsymbol{k} \cdot \boldsymbol{k} = 0$, according to Maxwell's equations [equation (22.30)]. Therefore it lies in a surface of constant phase:

$$\begin{pmatrix}\text{number of surfaces}\\ \text{pierced by } \boldsymbol{k}\end{pmatrix} = \langle \boldsymbol{d}\theta, \boldsymbol{k}\rangle = \langle \tilde{\boldsymbol{k}}, \boldsymbol{k}\rangle = \boldsymbol{k}\cdot\boldsymbol{k} = 0.$$

But not only does it lie in a surface of constant phase; it is also perpendicular to that surface! Any vector $\boldsymbol{v}$ in that surface must satisfy $\boldsymbol{k}\cdot\boldsymbol{v} = \langle \tilde{\boldsymbol{k}}, \boldsymbol{v}\rangle = \langle \boldsymbol{d}\theta, \boldsymbol{v}\rangle = 0$ because it pierces no surfaces.

Geometric optics assumes that the reduced wavelength $\lambda\!\!\!^{-}$, as measured in a typical local Lorentz frame, is small compared to the scale $\mathcal{L}$ of inhomogeneities in the wave train and small compared to the radius of curvature of spacetime, $\mathcal{R}$. Thus, over regions much larger than $\lambda\!\!\!^{-}$ but smaller than $\mathcal{L}$ or $\mathcal{R}$, the waves are plane-fronted and monochromatic, and there exist Lorentz reference frames (Riemann normal coordinates). In one of these "extended" local Lorentz frames, the phase must be

$$\theta = k_\alpha x^\alpha + \text{constant};$$

no other expression will yield $\boldsymbol{\nabla}\theta = \boldsymbol{k}$. The corresponding vector potential [equation (22.25)] will be

$$A^\mu = \Re\{a^\mu \exp[i(\boldsymbol{k}\cdot\boldsymbol{x} - k^0 t)]\} + (\text{"post-geometric-optics corrections"});$$

hence,

$$\begin{gathered} k^0 = 2\pi/(\text{period of wave}) = 2\pi\nu = \omega \equiv (\text{angular frequency}), \\ |\boldsymbol{k}| = 2\pi/(\text{wavelength of wave}) = 1/\lambdabar = \omega, \\ \boldsymbol{k} \text{ points along direction of propagation of wave.} \end{gathered}$$

At each event in spacetime there is a wave vector; and these wave vectors, tacked end-on-end, form a family of curves—the "light rays" or simply "rays"—whose tangent vector is $\boldsymbol{k}$. The rays, like their tangent vector, lie both in and perpendicular to the surfaces of constant phase.

The affine parameter λ of a ray (not to be confused with wavelength $= 2\pi\lambdabar$) satisfies $\boldsymbol{k} = d/d\lambda$; therefore it is given by

$$\lambda = t/k^0 + \text{constant} = t/\omega + \text{constant},$$

where t is proper time along the ray as measured, not by the ray itself (its proper time is zero!), but by the local Lorentz observer who sees angular frequency ω. Thus, while ω is a frame-dependent quantity and t is also a frame-dependent quantity, their quotient t/ω when measured along the ray (*not off the ray*) is the frame-independent affine parameter. For a particle it is possible and natural to identify the affine parameter λ with proper time τ. For a light ray this identification is unnatural and impossible. The lapse of proper time along the ray is identically zero. The springing up of λ to take the place of the vanished τ gives one a tool to do what one might not have suspected to be possible. Given a light ray shot out at event $\mathscr{A}$ and passing through event $\mathscr{B}$, one can give a third event $\mathscr{C}$ along the same null world line that is twice as "far" from $\mathscr{A}$ as $\mathscr{B}$ is "far," in a new sense of "far" that has nothing whatever directly to do with proper time (zero!), but is defined by equal increments of the affine parameter ($\lambda_{\mathscr{C}} - \lambda_{\mathscr{B}} = \lambda_{\mathscr{B}} - \lambda_{\mathscr{A}}$). The "affine parameter" has a meaning for any null geodesic analyzed even in isolation. In this respect, it is to be distinguished from the so-called "luminosity distance" which is sometimes introduced in dealing with the propagation of radiation through curved spacetime, and which is defined by the spreading apart of two or more light rays coming from a common source.

Maxwell's equations as explored in the text [equation (22.28′)] guarantee that the complex polarization vector $\boldsymbol{f}$ is perpendicular to the wave vector $\boldsymbol{k}$ and that, therefore, it lies in a surface of constant phase (see drawing). Intuition into the polarization vector is developed in exercise 22.12.

$$0\left(\frac{1}{\varepsilon^2}\right): \qquad k^\beta k_\beta a^\alpha = 0$$

$$\Longrightarrow \boldsymbol{k}\cdot\boldsymbol{k} = 0 \text{ (wave vector is null);} \tag{22.30}$$

$$0\left(\frac{1}{\varepsilon}\right): \qquad \underbrace{k^\beta k_\beta}_{[=0]} b^\alpha - 2i\left(k^\beta a^\alpha{}_{;\beta} + \frac{1}{2}k^\beta{}_{;\beta}a^\alpha\right) = 0$$

$$\Longrightarrow \boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{a} = -\frac{1}{2}(\boldsymbol{\nabla}\cdot\boldsymbol{k})\boldsymbol{a} \text{ (propagation equation for vector amplitude).} \tag{22.31}$$

These equations (22.30, 22.31) together with equation (22.28) are the basis from which all subsequent results will follow. As a first consequence, one can obtain the geodesic law from equation (22.30). Form the gradient of $\boldsymbol{k}\cdot\boldsymbol{k} = 0$,

$$0 = (k^\beta k_\beta)_{;\alpha} = 2k^\beta k_{\beta;\alpha},$$

and use the fact that $k_\beta \equiv \theta_{,\beta}$ is the gradient of a *scalar* to interchange indices, $\theta_{;\beta\alpha} = \theta_{;\alpha\beta}$ or

$$0 = k^\beta k_{\beta;\alpha} = k^\beta k_{\alpha;\beta}.$$

The main laws of geometric optics:

The result is

$$\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{k} = 0 \text{ (propagation equation for wave vector).} \tag{22.32}$$

(1) Light rays are null geodesics

Notice that this is the geodesic equation! Combined with equation (22.30), it is the statement, derived from Maxwell's equations in curved spacetime, that *light rays are null geodesics,* the first main result of geometric optics.

Turn now from the propagation vector $\boldsymbol{k} = \boldsymbol{\nabla}\theta$ to the wave amplitude $\boldsymbol{a} = a\boldsymbol{f}$, and obtain separate equations for the magnitude a and polarization $\boldsymbol{f}$. Use equation (22.31) to compute

$$\begin{aligned} 2a\,\partial_{\boldsymbol{k}}a = 2a\,\boldsymbol{\nabla}_{\boldsymbol{k}}a = \boldsymbol{\nabla}_{\boldsymbol{k}}a^2 &= \boldsymbol{\nabla}_{\boldsymbol{k}}(\boldsymbol{a}\cdot\bar{\boldsymbol{a}}) = \bar{\boldsymbol{a}}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{a} + \boldsymbol{a}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}}\bar{\boldsymbol{a}} \\ &= -\frac{1}{2}(\boldsymbol{\nabla}\cdot\boldsymbol{k})(\bar{\boldsymbol{a}}\cdot\boldsymbol{a} + \boldsymbol{a}\cdot\bar{\boldsymbol{a}}) = -a^2\,\boldsymbol{\nabla}\cdot\boldsymbol{k}; \end{aligned}$$

so

$$\partial_{\boldsymbol{k}}a = -\frac{1}{2}(\boldsymbol{\nabla}\cdot\boldsymbol{k})a \text{ (propagation equation for scalar amplitude).} \tag{22.33}$$

Next write $\boldsymbol{a} = a\boldsymbol{f}$ in equation (22.31) to obtain

$$0 = \boldsymbol{\nabla}_{\boldsymbol{k}}(a\boldsymbol{f}) + \frac{1}{2}(\boldsymbol{\nabla}\cdot\boldsymbol{k})a\boldsymbol{f} = a\,\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{f} + \boldsymbol{f}\left[\boldsymbol{\nabla}_{\boldsymbol{k}}a + \frac{1}{2}(\boldsymbol{\nabla}\cdot\boldsymbol{k})a\right] = a\,\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{f}$$

or

$$\nabla_{\boldsymbol{k}} \boldsymbol{f} = 0 \ (\textit{propagation equation for polarization vector}). \tag{22.34}$$

This together with equation (22.28′), constitutes the second main result of geometric optics, that *the polarization vector is perpendicular to the rays and is parallel-propagated along the rays*. It is now possible to see that these results, derived from equations (22.30) and (22.31) are consistent with the gauge condition (22.28). The vectors $\boldsymbol{k}$ and $\boldsymbol{f}$, specified at one point, are fixed along the entire ray by their propagation equations. But because both propagation equations are parallel-transport laws, the conditions $\boldsymbol{k} \cdot \boldsymbol{k} = 0$, $\boldsymbol{f} \cdot \bar{\boldsymbol{f}} = 1$, and $\boldsymbol{k} \cdot \boldsymbol{f} = 0$, once imposed on the vectors at one point, will be satisfied along the entire ray.

(2) polarization vector is perpendicular to ray and is parallel propagated along ray

The equation (22.33) for the scalar amplitude can be reformulated as a conservation law. Since $\partial_{\boldsymbol{k}} \equiv (\boldsymbol{k} \cdot \nabla)$, one rewrites the equation as $(\boldsymbol{k} \cdot \nabla) a^2 + a^2 \nabla \cdot \boldsymbol{k} = 0$, or

$$\nabla \cdot (a^2 \boldsymbol{k}) = 0. \tag{22.35}$$

(3) conservation of "photon number"

Consequently the vector $a^2\boldsymbol{k}$ is a "conserved current," and the integral $\int a^2 k^\mu \, d^3\Sigma_\mu$ has a fixed, unchanging value for each 3-volume cutting a given tube formed of light rays. (The tube must be so formed of rays that an integral of $a^2\boldsymbol{k}$ over the walls of the tube will give zero.) What is conserved? To remain purely classical, one could say it is the "number of light rays" and call a^2k^0 the "density of light rays" on an $x^0 =$ constant hypersurface. But the proper correspondence and more concrete physical interpretation make one prefer to call equation (22.35) *the law of conservation of photon number*. It is the third main result of geometric optics. Photon number, of course, is not always conserved; it is an adiabatic invariant, a quantity that is not changed by influences (e.g., spacetime curvature, $\sim 1/\mathcal{R}^2$) which change slowly ($\mathcal{R} \gg \lambda\!\!\!^{-}$) compared to the photon frequency.

Box 22.4 summarizes the above equations of geometric optics, along with others derived in the exercises.

EXERCISES

Exercise 22.11. ELECTROMAGNETIC FIELD AND STRESS ENERGY

Derive the equations given in part D of Box 22.4 for $\boldsymbol{F}$, $\boldsymbol{E}$, $\boldsymbol{B}$, and $\boldsymbol{T}$.

Exercise 22.12. POLARIZATION

At an event $\mathscr{P}_0$ through which geometric-optics waves are passing, introduce a local Lorentz frame with z-axis along the direction of propagation. Then $\boldsymbol{k} = \omega(\boldsymbol{e}_0 + \boldsymbol{e}_z)$. Since the polarization vector is orthogonal to $\boldsymbol{k}$, it is $\boldsymbol{f} = f^0(\boldsymbol{e}_0 + \boldsymbol{e}_z) + f^1\boldsymbol{e}_x + f^2\boldsymbol{e}_y$; and since $\boldsymbol{f} \cdot \bar{\boldsymbol{f}} = 1$, it has $|f^1|^2 + |f^2|^2 = 1$.

(a) Show that the component f^0 of the polarization vector has no influence on the electric and magnetic fields measured in the given frame; i.e., show that one can add a multiple of $\boldsymbol{k}$ to $\boldsymbol{f}$ without affecting any physical measurements.

(*continued on page 581*)

Box 22.4 GEOMETRIC OPTICS IN CURVED SPACETIME (Summary of Results Derived in Text and Exercises)

A. Geometric Optics Assumption

Electromagnetic waves propagating in a source-free region of spacetime are locally plane-fronted and monochromatic (reduced wavelength $\lambdabar \ll$ scale $\mathcal{L}$ over which amplitude, wavelength, or polarization vary; and $\lambdabar \ll \mathcal{R}$ = mean radius of curvature of spacetime).

B. Rays, Phase, and Wave Vector (see Box 22.3)

Everything (amplitude, polarization, energy, etc.) is transported along *rays;* and the quantities on one ray do not influence the quantities on any other ray.

The rays are null geodesics of curved spacetime, with tangent vectors ("*wave vectors*") $\boldsymbol{k}$:

$$\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{k} = 0.$$

The rays both lie in and are perpendicular to surfaces of constant phase, $\theta =$ const.; and their tangent vectors are the gradient of θ:

$$\boldsymbol{k} = \boldsymbol{\nabla}\theta.$$

In a local Lorentz frame, k^0 is the "angular frequency" and $k^0/2\pi$ is the ordinary frequency of the waves, and

$$\boldsymbol{n} = \boldsymbol{k}/k^0$$

is a unit 3-vector pointing along their direction of propagation.

C. Amplitude and Polarization Vector

The waves are characterized by a real amplitude a and a complex polarization vector $\boldsymbol{f}$ of unit length, $\boldsymbol{f} \cdot \bar{\boldsymbol{f}} = 1$. (Of the fundamental quantities θ, $\boldsymbol{k}$, a, $\boldsymbol{f}$, all are real except $\boldsymbol{f}$. See exercise 22.12 for deeper understanding of $\boldsymbol{f}$.)

The polarization vector is everywhere orthogonal to the rays, $\boldsymbol{k} \cdot \boldsymbol{f} = 0$; and is parallel-transported along them, $\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{f} = 0$.

The propagation law for the amplitude is

$$\partial_{\boldsymbol{k}} a = -\frac{1}{2}(\boldsymbol{\nabla} \cdot \boldsymbol{k})a.$$

This propagation law is equivalent to a *law of conservation of photons* (classically: of rays); $a^2\boldsymbol{k}$ is the "conserved current" satisfying $\boldsymbol{\nabla}\cdot(a^2\boldsymbol{k}) = 0$; and $(8\pi\hbar)^{-1}\int a^2k^0\sqrt{|g|}\,d^3x$ is the number of photons (rays) in the 3-volume of integration on any $x^0 =$ constant hypersurface, and is constant as this volume is carried along the rays.

The propagation law holds separately on each hypersurface of constant phase. There it can be interpreted as conservation of a $a^2\mathcal{A}$, where $\mathcal{A}$ is a two-dimensional cross-sectional area of a pulse of photons or rays. See exercise 22.13.

D. Vector Potential, Electromagnetic Field, and Stress-Energy-Momentum

At any event the vector potential in Lorentz gauge is

$$\boldsymbol{A} = \Re\{ae^{i\theta}\boldsymbol{f}\},$$

where $\Re$ denotes the real part.

The electromagnetic field tensor is orthogonal to the rays, $\boldsymbol{F}\cdot\boldsymbol{k} = 0$, and is given by

$$\boldsymbol{F} = \Re\{iae^{i\theta}\boldsymbol{k}\wedge\boldsymbol{f}\}.$$

The corresponding electric and magnetic fields in any local Lorentz frame are

$$\boldsymbol{E} = \Re\{iak^0e^{i\theta}(\text{projection of } \boldsymbol{f} \text{ perpendicular to } \boldsymbol{k})\},$$
$$\boldsymbol{B} = \boldsymbol{n}\times\boldsymbol{E}, \text{ where } \boldsymbol{n}\equiv\boldsymbol{k}/k^0.$$

The stress-energy tensor, averaged over a wavelength, is

$$\boldsymbol{T} = (1/8\pi)a^2\boldsymbol{k}\otimes\boldsymbol{k},$$

corresponding to an energy density in a local Lorentz frame of

$$T^{00} = (1/8\pi)(ak^0)^2$$

and an energy flux of

$$T^{0j} = T^{00}n^j,$$

so that energy flows along the rays (in $\boldsymbol{n} = \boldsymbol{k}/k^0$ direction) with the speed of light. This is identical with the stress-energy tensor that would be produced by a beam of photons with 4-momenta $\boldsymbol{p} = \hbar\boldsymbol{k}$.

Conservation of energy-momentum $\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$ follows from the ray conservation law $\boldsymbol{\nabla}\cdot(a^2\boldsymbol{k}) = 0$ and the geodesic law $\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{k}\equiv(\boldsymbol{k}\cdot\boldsymbol{\nabla})\boldsymbol{k} = 0$:

$$8\pi\,\boldsymbol{\nabla}\cdot\boldsymbol{T} = \boldsymbol{\nabla}\cdot(a^2\boldsymbol{k}\otimes\boldsymbol{k}) = [\boldsymbol{\nabla}\cdot(a^2\boldsymbol{k})]\boldsymbol{k} + a^2(\boldsymbol{k}\cdot\boldsymbol{\nabla})\boldsymbol{k} = 0.$$

Box 22.4 ***(continued)***

The adiabatic (geometric optics) invariant "ray number" a^2k^0 or "photon number" $(8\pi\hbar)^{-1}a^2k^0$ in a unit volume is proportional to the energy, $(8\pi)^{-1}a^2(k^0)^2$, divided by the frequency, k^0—corresponding exactly to the harmonic oscillator adiabatic invariant E/ω [Einstein (1912), Ehrenfest (1916), Landau and Lifshitz (1960)].

E. Photon Reinterpretation of Geometric Optics

The laws of geometric optics can be reinterpreted as follows. This reinterpretation becomes a foundation of the standard quantum theory of the electromagnetic field (see, e.g., Chapters 1 and 13 of Baym (1969)]; and the classical limit of that quantum theory is standard Maxwell electrodynamics.

Photons are particles of zero rest mass that move along null geodesics of spacetime (the null rays).

The 4-momentum of a photon is related to the tangent vector of the null ray (wave vector) by $\boldsymbol{p} = \hbar \boldsymbol{k}$. A renormalization of the affine parameter,

$$(\text{new parameter}) = (1/\hbar) \times (\text{old parameter}),$$

makes $\boldsymbol{p}$ the tangent vector to the ray.

Each photon possesses a polarization vector, $\boldsymbol{f}$, which is orthogonal to its 4-momentum ($\boldsymbol{p} \cdot \boldsymbol{f} = 0$), and which it parallel-transports along its geodesic world line ($\boldsymbol{\nabla}_{\boldsymbol{p}}\boldsymbol{f} = 0$).

A swarm of photons, all with nearly the same 4-momentum $\boldsymbol{p}$ and polarization vector $\boldsymbol{f}$ (as compared by parallel transport), make up a classical electromagnetic wave. The scalar amplitude a of the wave is determined by equating the stress-energy tensor of the wave

$$\boldsymbol{T} = \frac{1}{8\pi} a^2 \boldsymbol{k} \otimes \boldsymbol{k} = \frac{1}{8\pi}\left(\frac{a}{\hbar}\right)^2 \boldsymbol{p} \otimes \boldsymbol{p}$$

to the stress-energy tensor of a swarm of photons with number-flux vector $\boldsymbol{S}$,

$$\boldsymbol{T} = \boldsymbol{p} \otimes \boldsymbol{S}$$

[see equation (5.18)]. The result:

$$\boldsymbol{S} = \frac{1}{8\pi}\left(\frac{a}{\hbar}\right)^2 \boldsymbol{p} = \frac{1}{8\pi\hbar} a^2 \boldsymbol{k}$$

or, in any local Lorentz frame,

$$a = (8\pi\hbar^2 S^0/p^0)^{1/2} = (8\pi)^{1/2}\hbar \left(\frac{\text{number density of photons}}{\text{energy of one photon}}\right)^{1/2}.$$

(b) Show that the following polarization vectors correspond to the types of polarization listed:

$\boldsymbol{f} = \boldsymbol{e}_x$, linear polarization in x direction;

$\boldsymbol{f} = \boldsymbol{e}_y$, linear polarization in y direction;

$\boldsymbol{f} = \frac{1}{\sqrt{2}}(\boldsymbol{e}_x + i\boldsymbol{e}_y)$, righthand circular polarization;

$\boldsymbol{f} = \frac{1}{\sqrt{2}}(\boldsymbol{e}_x - i\boldsymbol{e}_y)$, lefthand circular polarization;

$\boldsymbol{f} = \alpha\boldsymbol{e}_x + i(1 - \alpha^2)^{1/2}\boldsymbol{e}_y$, righthand elliptical polarization.

(c) Show that the type of polarization (linear; circular; elliptical with given eccentricity of ellipse) is the same as viewed in any local Lorentz frame at any event along a given ray. [*Hint:* Use pictures and abstract calculations rather than Lorentz transformations and component calculations.]

Exercise 22.13. THE AREA OF A BUNDLE OF RAYS

Write equation (22.31) in a coordinate system in which one of the coordinates is chosen to be $x^0 = \theta$, the phase (a retarded time coordinate).

(a) Show that $g^{00} = 0$ and that no derivatives $\partial/\partial\theta$ appear in equation (22.33); so propagation of a can be described within a single $\theta =$ constant hypersurface.

(b) Perform the following construction (see Figure 22.1). Pick a ray $\mathscr{C}_0$ along which a is to be propagated. Pick a bundle of rays, with two-dimensional cross section, that (i) all lie in the same constant-phase surface as $\mathscr{C}_0$, and (ii) surround $\mathscr{C}_0$. (The surface is three-di-

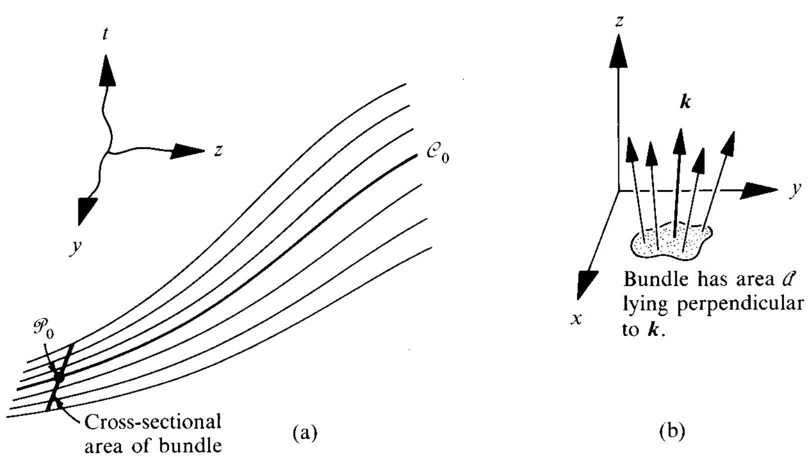


Figure 22.1.
Geometric optics for a bundle of rays with two-dimensional cross section, all lying in a surface of constant phase, $\theta =$ const. Sketch (a) shows the bundle, surrounding a central ray $\mathscr{C}_0$, in a spacetime diagram with one spatial dimension suppressed. Sketch (b) shows the bundle as viewed on a slice of simultaneity in a local Lorentz frame at the event $\mathscr{P}_0$. Slicing the bundle turns each ray into a "photon"; so the bundle becomes a two-dimensional surface filled with photons. The area $\mathscr{A}$ of this photon-filled surface obeys the following laws (see exercises 22.13 and 22.14): (1) $\mathscr{A}$ is independent of the choice of Lorentz frame; it depends only on location $\mathscr{P}_0$ along the ray $\mathscr{C}_0$. (2) The amplitude a of the waves satisfies

$$\mathscr{A}a^2 = \text{constant all along the ray } \mathscr{C}_0$$

("*conservation of photon flux*"). (3) $\mathscr{A}$ obeys the "*propagation equation*" (22.36).

mensional, so any bundle filling it has a two-dimensional cross section.) At any event $\mathscr{P}_0$, in any local Lorentz frame there, on a "slice of simultaneity" $x^0 =$ constant, measure the cross-sectional area $\mathcal{A}$ of the bundle. (Note: the area being measured is perpendicular to $\boldsymbol{k}$ in the three-dimensional Euclidean sense; it can be thought of as the region occupied momentarily by a group of photons propagating along, side by side, in the $\boldsymbol{k}$ direction.) Show that the area $\mathcal{A}$ is the same, at a given event $\mathscr{P}_0$, *regardless of what Lorentz frame is used to measure it; but the area changes from point to point along the ray* $\mathscr{C}_0$ *as a result of the rays' divergence away from each other or convergence toward each other:*

$$\partial_{\boldsymbol{k}}\mathcal{A} = (\boldsymbol{\nabla}\cdot\boldsymbol{k})\mathcal{A}. \tag{22.36}$$

Then show that $\mathcal{A}a^2$ is a constant everywhere along the ray $\mathscr{C}_0$ ("conservation of photon flux"). [Hints: (i) Any vector $\boldsymbol{\xi}$ connecting adjacent rays in the bundle is perpendicular to $\boldsymbol{k}$, because $\boldsymbol{\xi}$ lies in a surface of constant θ and $\boldsymbol{k}\cdot\boldsymbol{\xi} = \langle\tilde{\boldsymbol{k}}, \boldsymbol{\xi}\rangle = \langle\boldsymbol{d}\theta, \boldsymbol{\xi}\rangle =$ (change in θ along $\boldsymbol{\xi}$) $= 0$. (ii) Consider, for simplicity, a bundle with rectangular cross section as seen in a specific local Lorentz frame at a specific event $\mathscr{P}_0$ [edge vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ with $\boldsymbol{v}\cdot\boldsymbol{w} = 0$ (edges perpendicular) and $\boldsymbol{v}\cdot\boldsymbol{e}_0 = \boldsymbol{w}\cdot\boldsymbol{e}_0 = 0$ (edges in surface of constant time) and $\boldsymbol{v}\cdot\boldsymbol{k} = \boldsymbol{w}\cdot\boldsymbol{k} = 0$ (since edge vectors connect adjacent rays of the bundle)]. Show pictorially that in any other Lorentz frame at $\mathscr{P}_0$, the edge vectors are $\boldsymbol{v}' = \boldsymbol{v} + \alpha\boldsymbol{k}$ and $\boldsymbol{w}' = \boldsymbol{w} + \beta\boldsymbol{k}$ for some α and β. Conclude that in all Lorentz frames at $\mathscr{P}_0$ the cross section has identical shape and identical area, and is spatially perpendicular to the direction of propagation ($\boldsymbol{k}\cdot\boldsymbol{v} = \boldsymbol{k}\cdot\boldsymbol{w} = 0$). (iii) By a calculation in a local Lorentz frame show that $\partial_{\boldsymbol{k}}\mathcal{A} = (\boldsymbol{\nabla}\cdot\boldsymbol{k})\mathcal{A}$. (iv) Conclude from $\partial_{\boldsymbol{k}}a = -\frac{1}{2}(\boldsymbol{\nabla}\cdot\boldsymbol{k})a$ that $\partial_{\boldsymbol{k}}(\mathcal{A}a^2) = 0$.]

Exercise 22.14. FOCUSING THEOREM

The cross-sectional area $\mathcal{A}$ of a bundle of rays all lying in the same surface of constant phase changes along the central ray of the bundle at the rate (22.36) (see Figure 22.1).

(a) Derive the following equation ("*focusing equation*") for the second derivative of $\mathcal{A}^{1/2}$:

$$\frac{d^2\mathcal{A}^{1/2}}{d\lambda^2} = -\left(|\sigma|^2 + \frac{1}{2}R_{\alpha\beta}k^\alpha k^\beta\right)\mathcal{A}^{1/2}, \tag{22.37}$$

where λ is affine parameter along the central ray ($\boldsymbol{k} = d/d\lambda$), and the "magnitude of the shear of the rays", $|\sigma|$, is defined by the equation

$$|\sigma|^2 \equiv \frac{1}{2}k_{\alpha;\beta}k^{\alpha;\beta} - \frac{1}{4}(k^\mu{}_{;\mu})^2. \tag{22.38}$$

[Hint: This is a vigorous exercise in index manipulations. The key equations needed in the manipulations are $\mathcal{A}_{,\alpha}k^\alpha = (k^\alpha{}_{;\alpha})\mathcal{A}$ [equation (22.36)]; $k^\alpha{}_{;\beta}k^\beta = 0$ [geodesic equation (22.32) for rays]; $k_{\alpha;\beta} = k_{\beta;\alpha}$ [which follows from $k_\alpha \equiv \theta_{,\alpha}$]; and the rule (16.6c) for interchanging covariant derivatives of a vector.]

(b) Show that, in a local Lorentz frame where $\boldsymbol{k} = \omega(\boldsymbol{e}_t + \boldsymbol{e}_z)$ at the origin,

$$|\sigma|^2 = \frac{1}{4}(k_{x,x} - k_{y,y})^2 + (k_{x,y})^2. \tag{22.39}$$

Thus, $|\sigma|^2$ is nonnegative, which justifies the use of the absolute value sign.

(c) *Discussion:* The quantity $|\sigma|$ is called the *shear* of the bundle of rays because it measures the extent to which neighboring rays are sliding past each other [see, e.g., Sachs (1964)]. Hence, the focusing equation (22.37) says that shear focuses a bundle of rays (makes $d^2\mathcal{A}^{1/2}/d\lambda^2 < 0$); and spacetime curvature also focuses it if $R_{\alpha\beta}k^\alpha k^\beta > 0$, but defocuses it if $R_{\alpha\beta}k^\alpha k^\beta < 0$. (When a bundle of toothpicks, originally circular in cross section, is squeezed into an elliptic cross section, it is sheared.)

(d) Assume that the energy density $T_{\hat{0}\hat{0}}$, as measured by any observer anywhere in spacetime, is nonnegative. By combining the focusing equation (22.37) with the Einstein field equation, conclude that

$$\frac{d^2 \mathcal{A}^{1/2}}{d\lambda^2} \leq 0 \begin{pmatrix} \text{for any bundle of rays, all in the same} \\ \text{surface of constant phase, anywhere in} \\ \text{spacetime} \end{pmatrix} \tag{22.40}$$

(*focusing theorem*). This theorem plays a crucial role in black-hole physics (§34.5) and in the theory of singularities (§34.6).

§22.6. KINETIC THEORY IN CURVED SPACETIME*

The stars in a galaxy wander through spacetime, each on its own geodesic world line, each helping to produce the spacetime curvature felt by all the others. Photons, left over from the hot phases of the big bang, bathe the Earth, bringing with themselves data on the homogeneity and isotropy of the universe. Theoretical analyses of these and many other problems are unmanageable, if they attempt to keep track of the motion of every single star or photon. But a statistical description gives accurate results and is powerful. Moreover, for most problems in astrophysics and cosmology, the simplest of statistical descriptions—one ignoring collisions—is adequate. Usually collisions are unimportant for the large-scale behavior of a system (e.g.; a galaxy), or they are so important that a fluid description is possible (e.g., in a stellar interior).

Consider, then, a swarm of particles (stars, or photons, or black holes, or . . .) that move through spacetime on geodesic world lines, without colliding. Assume, for simplicity, that the particles all have the same rest mass. Then all information of a statistical nature about the particles can be incorporated into a single function, the "*distribution function*" or "*number density in phase space*", $\mathcal{N}$.

Volume in phase space for a group of identical particles

Define $\mathcal{N}$ in terms of measurements made by a specific local Lorentz observer at a specific event $\mathcal{P}_0$ in curved spacetime. Give the observer a box with 3-volume $\mathcal{V}_x$ (and with imaginary walls). Ask the observer to count how many particles, N, are inside the box *and* have local-Lorentz momentum components p^j in the range

$$P^j - \frac{1}{2}\Delta p^j < p^j < P^j + \frac{1}{2}\Delta p^j.$$

(He can ignore the particle energies p^0; since all particles have the same rest mass m, energy

$$p^0 = (m^2 + \boldsymbol{p}^2)^{1/2}$$

*For more detailed and sophisticated treatments of this topic, see, e.g., Tauber and Weinberg (1961), and Lindquist (1966), Marle (1969), Ehlers (1971), Stewart (1971), Israel (1972), and references cited therein. Ehlers (1971) is a particularly good introductory review article.

is fixed uniquely by momentum.) The volume in momentum space occupied by the N particles is $\mathcal{V}_p = \Delta p^x \, \Delta p^y \, \Delta p^z$; and *the volume in phase space is*

$$\mathcal{V} \equiv \mathcal{V}_x \mathcal{V}_p. \tag{22.41}$$

Lorentz invariance of volume in phase space

Other observers at $\mathcal{P}_0$, moving relative to the first, will disagree on how much spatial volume $\mathcal{V}_x$ and how much momentum volume $\mathcal{V}_p$ these same N particles occupy:

$$\mathcal{V}_x \text{ and } \mathcal{V}_p \text{ depend on the choice of Lorentz frame.} \tag{22.42}$$

However, all observers will agree on the value of the product $\mathcal{V} \equiv \mathcal{V}_x \mathcal{V}_p$ ("volume in phase space"):

> The phase-space volume $\mathcal{V}$ occupied by a given set of N identical particles at a given event in spacetime is independent of the local Lorentz frame in which it is measured. (22.43)

(See Box 22.5 for proof.) Moreover, as the same N particles move through spacetime along their geodesic world lines (and through momentum space), the volume $\mathcal{V}$ they span in phase space remains constant:

Liouville's theorem (conservation of volume in phase space)

> The $\mathcal{V}$ occupied by a given swarm of N particles is independent of location along the world line of the swarm ("*Liouville's theorem in curved spacetime*"). (22.44)

(See Box 22.6 for proof.)

Number density in phase space (distribution function)

More convenient for applications than the volume $\mathcal{V}$ in phase space occupied by a given set of N particles is the "*number density in phase space*" ("*distribution function*") in the neighborhood of one of these particles:

$$\mathcal{N} \equiv N/\mathcal{V}. \tag{22.45}$$

On what does this number density depend? It depends on the location in spacetime, $\mathcal{P}$, at which the measurements are made. It also depends on the 4-momentum $\boldsymbol{p}$ of the particle in whose neighborhood the measurements are made. But because the particles all have the same rest mass, $\boldsymbol{p}$ cannot take on any and every value in the tangent space at $\mathcal{P}$. Rather, $\boldsymbol{p}$ is confined to the "forward mass hyperboloid" at $\mathcal{P}$:

$$\boldsymbol{p}^2 = m^2; \qquad \boldsymbol{p} \text{ lies inside future light cone.}$$

Thus,

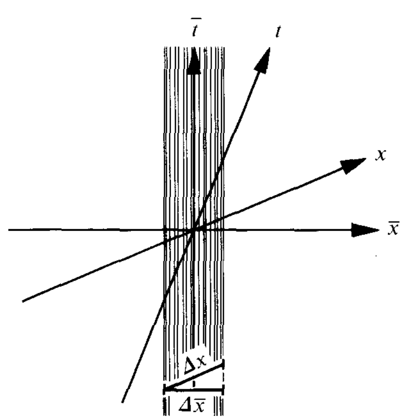
$$\mathcal{N} = \mathcal{N}\left[\begin{pmatrix}\text{location, } \mathcal{P}, \\ \text{in spacetime}\end{pmatrix}, \begin{pmatrix}\text{4-momentum } \boldsymbol{p}\text{, which must lie} \\ \text{on the forward mass hyperboloid} \\ \text{of the tangent space at } \mathcal{P}\end{pmatrix}\right]. \tag{22.46}$$

Pick some one particle in the swarm, with geodesic world line $\mathcal{P}(\lambda)$ [λ = (affine parameter) = (proper time, if particle has finite rest mass)], and with 4-momentum

Box 22.5 VOLUME IN PHASE SPACE

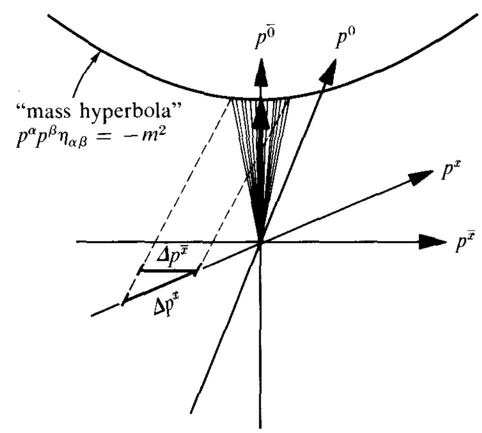
A. For Swarm of Identical Particles with Nonzero Rest Mass

Pick an event $\mathscr{P}_0$, through which passes a particle named "John" with a 4-momentum named "$\boldsymbol{P}$". In John's local Lorentz rest frame at $\mathscr{P}_0$ ("barred frame", $\bar{\mathcal{S}}$), select a small 3-volume, $\mathscr{V}_{\bar{x}} \equiv \Delta\bar{x}\,\Delta\bar{y}\,\Delta\bar{z}$, containing him. Also select a small "3-volume in momentum space," $\mathscr{V}_{\bar{p}} \equiv \Delta p^{\bar{x}}\,\Delta p^{\bar{y}}\,\Delta p^{\bar{z}}$ centered on John's momentum, which is $P^{\bar{x}} = P^{\bar{y}} = P^{\bar{z}} = 0$. Focus attention on all particles whose world lines pass through $\mathscr{V}_{\bar{x}}$ and which have momenta $p^{\bar{j}}$ in the range $\mathscr{V}_{\bar{p}}$ surrounding $P^{\bar{j}} = 0$.



Examine this bundle in another local Lorentz frame ("unbarred frame", $\mathcal{S}$) at $\mathscr{P}_0$, which moves with speed β relative to the rest frame. Orient axes so the relative motion of the frames is in the x and $\bar{x}$ directions. Then the space volume $\mathscr{V}_x$ occupied in the new frame has $\Delta y = \Delta\bar{y}$, $\Delta z = \Delta\bar{z}$ (no effect of motion on transverse directions), and $\Delta x = (1-\beta^2)^{1/2}\,\Delta\bar{x}$ (Lorentz contraction in longitudinal direction). Hence $\mathscr{V}_x = (1-\beta^2)^{1/2}\mathscr{V}_{\bar{x}}$ ("transformation law for space volumes") or, equivalently [since $P^0 = m/(1-\beta^2)^{1/2}$]:

$$P^0\mathscr{V}_x = m\mathscr{V}_{\bar{x}} = \begin{pmatrix}\text{constant, independent}\\ \text{of Lorentz frame}\end{pmatrix}.$$



A momentum-space diagram, analogous to the spacetime diagram, depicts the momentum spread for particles in the bundle, and shows that $\Delta p^x = \Delta p^{\bar{x}}/(1-\beta^2)^{1/2}$. The Lorentz transformation from $\bar{\mathcal{S}}$ to $\mathcal{S}$ leaves transverse components of momenta unaffected; so $\Delta p^y = \Delta p^{\bar{y}}$, $\Delta p^z = \Delta p^{\bar{z}}$. Hence $\mathscr{V}_p = \mathscr{V}_{\bar{p}}/(1-\beta^2)^{1/2}$ ("transformation law for momentum volumes"); or, equivalently

$$\frac{\mathscr{V}_p}{P^0} = \frac{\mathscr{V}_{\bar{p}}}{m} = \begin{pmatrix}\text{constant, independent}\\ \text{of Lorentz frame}\end{pmatrix}.$$

Although the spatial 3-volumes $\mathscr{V}_x$ and $\mathscr{V}_{\bar{x}}$ differ from one frame to another, and the momentum 3-volumes $\mathscr{V}_p$ and $\mathscr{V}_{\bar{p}}$ differ, the volume in six-dimensional phase space is Lorentz-invariant:

$$\mathscr{V} \equiv \mathscr{V}_{\bar{x}}\mathscr{V}_{\bar{p}} = \mathscr{V}_x\mathscr{V}_p.$$

It is a frame-independent, geometric object!

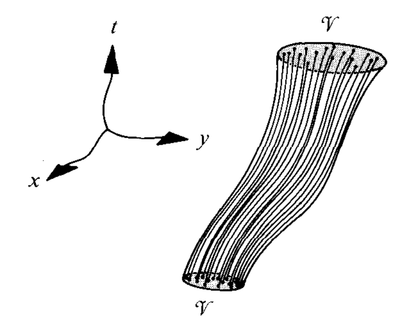
B. For Swarm of Identical Particles with Zero Rest Mass

Examine a sequence of systems, each with particles of smaller rest mass and of higher velocity relative to a laboratory. For every bundle of particles in each system, $P^0\mathscr{V}_x$, $\mathscr{V}_p/P^0$, and $\mathscr{V}_x\mathscr{V}_p$ are Lorentz-invariant. Hence, in the limit as $m \longrightarrow 0$, as $\beta \longrightarrow 1$, and as $P^0 = m/(1-\beta^2)^{1/2} \longrightarrow$ finite value (particles of zero rest mass moving with speed of light), $P^0\mathscr{V}_x$ and $\mathscr{V}_p/P^0$ and $\mathscr{V}_x\mathscr{V}_p$ are still Lorentz-invariant, geometric quantities.

Box 22.6 CONSERVATION OF VOLUME IN PHASE SPACE

Examine a very small bundle of identical particles that move through curved spacetime on neighboring geodesics. Measure the bundle's volume in phase space, $\mathcal{V}$ ($\mathcal{V} = \mathcal{V}_x \mathcal{V}_p$ in any local Lorentz frame), as a function of affine parameter λ along the central geodesic of the bundle. The following calculation shows that

$$d\mathcal{V}/d\lambda = 0 \qquad \left(\begin{array}{l}\text{“Liouville theorem in}\\ \text{curved spacetime”}\end{array}\right).$$



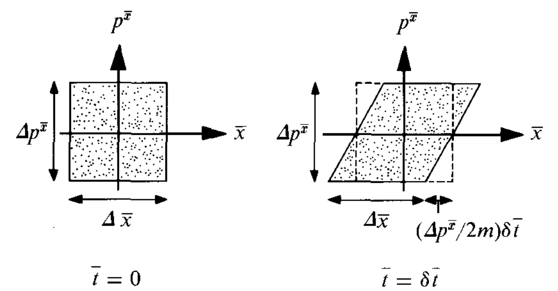
Proof for particles of finite rest mass: Examine particle motion during time interval $\delta\tau$, using local Lorentz rest frame of central particle. All velocities are small in this frame, so

$$p^{\bar{j}} = m dx^{\bar{j}}/d\bar{t}.$$

Hence (see pictures) the spreads in momentum and position conserve $\Delta\bar{x}\,\Delta p^{\bar{x}}$, $\Delta\bar{y}\,\Delta p^{\bar{y}}$, and $\Delta\bar{z}\,\Delta p^{\bar{z}}$; i.e.,

$$\frac{d\mathcal{V}}{d\tau} = \frac{\delta(\Delta\bar{x}\,\Delta\bar{y}\,\Delta\bar{z}\,\Delta p^{\bar{x}}\,\Delta p^{\bar{y}}\,\Delta p^{\bar{z}})}{\delta\bar{t}} = 0.$$

But $\tau = a\lambda + b$ for some arbitrary constants a and b; so $d\mathcal{V}/d\lambda = 0$.



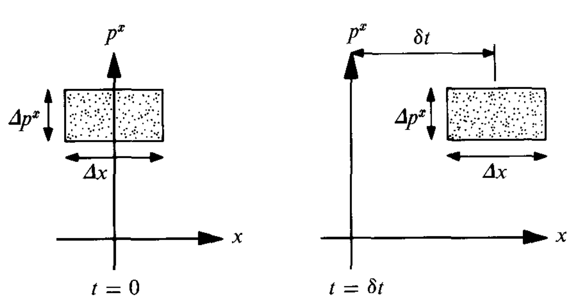
Each particle moves with speed $d\bar{x}/d\bar{t}$ proportional to height in diagram

$$d\bar{x}/d\bar{t} = p^{\bar{x}}/m,$$

and conserves its momentum, $dp^{\bar{x}}/d\bar{t} = 0$. Hence the region occupied by particles deforms, but maintains its area. Same is true for $(y - p^y)$ and $(z - p^z)$.

Proof for particles of zero rest mass. Examine particle motion in local Lorentz frame where central particle has $\boldsymbol{P} = P^0(\boldsymbol{e}_0 + \boldsymbol{e}_x)$. In this frame, all particles have $p^y \ll p^0$, $p^z \ll p^0$, $p^x = p^0 + O([p^y]^2/P^0) \approx P^0$. Since $p^\alpha = dx^\alpha/d\lambda$ for appropriate normalization of affine parameters (see Box 22.4), one can write $dx^j/dt = p^j/p^0$; i.e.,

$$\frac{dx}{dt} = 1 + O([p^y/P^0]^2 + [p^z/P^0]^2)$$
$$\approx 1,$$
$$\frac{dy}{dt} = \frac{p^y}{P^0}, \qquad \frac{dz}{dt} = \frac{p^z}{P^0}.$$



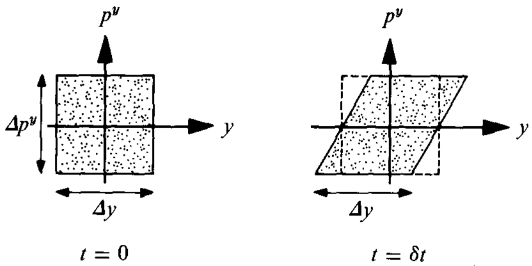
Each particle ("photon") moves with $dx/dt = 1$ and $dp^x/dt = 0$ in the local Lorentz frame. Area and shape of occupied region are preserved.

Hence (see pictures) $\Delta x\, \Delta p^x$, $\Delta y\, \Delta p^y$, and $\Delta z\, \Delta p^z$ are all conserved; and

$$\frac{d\mathcal{V}}{dt} = \frac{\delta(\Delta x\, \Delta y\, \Delta z\, \Delta p^x\, \Delta p^y\, \Delta p^z)}{\delta t} = 0.$$

But t and the affine parameter λ of central particle are related by $t = P^0\lambda$ [cf. equation (16.4)]; thus

$$d\mathcal{V}/d\lambda = 0.$$



Particle ("photon") speeds are proportional to height in diagram

$$dy/dt = p^y/P^0,$$

and $dp^y/dt = 0$. Hence, occupied region deforms but maintains its area. Same is true of $z - p^z$.

$\boldsymbol{p}(\lambda)$. Examine the density in phase space in this particle's neighborhood at each point along its world line:

$$\mathcal{N} = \mathcal{N}[\mathcal{P}(\lambda), \boldsymbol{p}(\lambda)].$$

Calculate $\mathcal{N}(\lambda)$ as follows: (1) Pick an initial event $\mathcal{P}(0)$ on the world line, and a phase-space volume $\mathcal{V}$ containing the particle. (2) Cover with red paint all the particles contained in $\mathcal{V}$ at $\mathcal{P}(0)$. (3) Watch the red particles move through spacetime alongside the initial particle. (4) As they move, the phase-space region they occupy changes shape extensively; but its volume $\mathcal{V}$ remains fixed (Liouville's theorem). Moreover, no particles can enter or leave that phase-space region (once in, always in; once out, always out; boundaries of phase-space region are attached to and move with the particles). (5) Hence, at any λ along the initial particle's world line, the particle is in a phase-space region of unchanged volume $\mathcal{V}$, unchanged number of particles N, and unchanged ratio $\mathcal{N} = N/\mathcal{V}$:

$$\frac{d\mathcal{N}[\mathcal{P}(\lambda), \boldsymbol{p}(\lambda)]}{d\lambda} = 0. \qquad (22.47)$$

Collisionless Boltzmann equation (kinetic equation)

This equation for the conservation of $\mathcal{N}$ along a particle's trajectory in phase space is called the "collisionless Boltzmann equation," or the "kinetic equation."

Photons provide an important application of the Boltzmann equation. But when discussing photons one usually does not think in terms of the number density in phase space. Rather, one speaks of the "specific intensity" I_ν of radiation at a given frequency ν, flowing in a given direction, $\boldsymbol{n}$, as measured in a specified local Lorentz frame:

$$I_\nu \equiv \frac{d(\text{energy})}{d(\text{time})\, d(\text{area})\, d(\text{frequency})\, d(\text{solid angle})}. \qquad (22.48)$$

Distribution function for photons expressed in terms of specific intensity, I_ν

(See Figure 22.2). A simple calculation in the local Lorentz frame reveals that

$$\mathcal{N} = h^{-4}(I_\nu/\nu^3), \tag{22.49}$$

Invariance and conservation of I_ν/ν^3

where h is Planck's constant (see Figure 22.2). Thus, if two different observers at the same or different events in spacetime look at the same photon (and neighboring photons) as it passes them, they will see different frequencies ν ("doppler shift," "cosmological red shift," "gravitational redshift"), and different specific intensities I_ν; but they will obtain identical values for the ratio I_ν/ν^3. Thus I_ν/ν^3, like $\mathcal{N}$, is invariant from observer to observer and from event to event along a given photon's world line.

EXERCISES

Exercise 22.15. INVERSE SQUARE LAW FOR FLUX

The *specific flux* of radiation entering a telescope from a given source is defined by

$$F_\nu = \int I_\nu \, d\Omega, \tag{22.50}$$

where integration is over the total solid angle (assumed $\ll 4\pi$) subtended by the source on the observer's sky. Use the Boltzmann equation (conservation of I_ν/ν^3) to show that $F_\nu \propto$ (distance from source)$^{-2}$ for observers who are all at rest relative to each other in flat spacetime.

Exercise 22.16. BRIGHTNESS OF THE SUN

Does the surface of the sun look any brighter to an astronaut standing on Mercury than to a student standing on Earth?

Exercise 22.17. BLACK BODY RADIATION

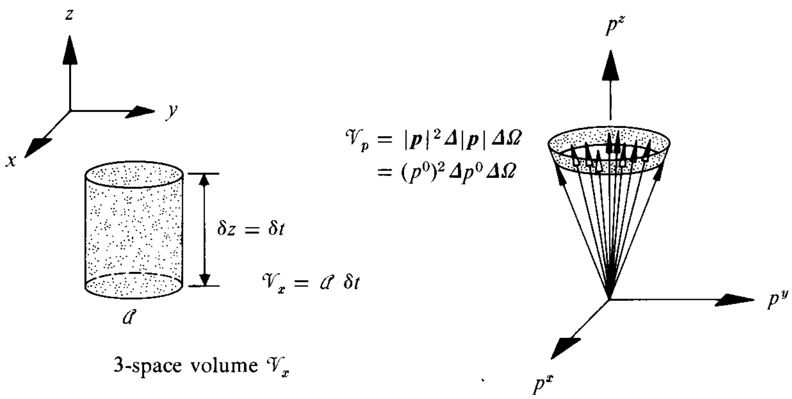
An "optically thick" source of black-body radiation (e.g., the surface of a star, or the hot matter filling the universe shortly after the big bang) emits photons isotropically with a specific intensity, as seen by an observer at rest near the source, given (Planck radiation law) by

$$I_\nu = \frac{2h\nu^3}{e^{h\nu/kT} - 1}. \tag{22.51}$$

Here T is the temperature of the source. Show that any observer, in any local Lorentz frame, anywhere in the universe, who examines this radiation as it flows past him, will also see a black-body spectrum. Show, further, that if he calculates a temperature by measuring the specific intensity I_ν at any one frequency, and if he calculates a temperature from the shape of the spectrum, those temperatures will agree. (Radiation remains black body rather than being "diluted" into "grey-body.") Finally, show that the temperature he measures is red-shifted by precisely the same factor as the frequency of any given photon is redshifted,

$$\frac{T_{\text{observed}}}{T_{\text{emitted}}} = \left(\frac{\nu_{\text{observed}}}{\nu_{\text{emitted}}}\right) \text{ for a given photon.} \tag{22.52}$$

[Note that the redshifts can be "Doppler" in origin, "cosmological" in origin, "gravitational" in origin, or some inseparable mixture. All that matters is the fact that the parallel-transport law for a photon's 4-momentum, $\boldsymbol{\nabla}_{\boldsymbol{p}} \boldsymbol{p} = 0$, guarantees that the redshift $\nu_{\text{observed}}/\nu_{\text{emitted}}$ is independent of frequency emitted.]



3-space volume $\mathcal{V}_x$

3-momentum volume, with direction of momentum vectors reversed for ease of visualization (telescope as an emitter, not a receiver!)

Figure 22.2.
Number density in phase space for photons, interpreted in terms of the specific intensity I_ν. An astronomer has a telescope with filter that admits only photons arriving from within a small solid angle $\Delta\Omega$ about the z-direction, and having energies between p^0 and $p^0 + \Delta p^0$. The collecting area, $\mathcal{A}$, of his telescope lies in the x, y-plane (perpendicular to the incoming photon beam). Let δN be the number of photons that cross the area $\mathcal{A}$ in a time interval δt. [All energies, areas, times, and lengths are measured in the orthonormal frame ("proper reference frame; §13.6) which the astronomer Fermi-Walker transports with himself along his (possibly accelerated) world line—or, equivalently, in a local Lorentz frame momentarily at rest with respect to the astronomer.] The δN photons, just before the time interval δt begins, lie in the cylinder of area $\mathcal{A}$ and height $\delta z = \delta t$ shown above. Their spatial 3-volume is thus $\mathcal{V}_x = \mathcal{A}\, \delta t$. Their momentum 3-volume is $\mathcal{V}_p = (p^0)^2\, \Delta p^0\, \Delta\Omega$ (see drawing). Hence, their number density in phase space is

$$\mathcal{N} = \frac{\delta N}{\mathcal{V}_x \mathcal{V}_p} = \frac{\delta N}{\mathcal{A}\, \delta t (p^0)^2 (\Delta p^0)\, \Delta\Omega} = \frac{\delta N}{h^3\, \mathcal{A}\, \delta t\, \nu^2\, \Delta\nu\, \Delta\Omega}$$

where ν is the photon frequency measured by the telescope ($p^0 = h\nu$).

The specific intensity of the photons, I_ν (a standard concept in astronomy), is the energy per unit area per unit time per unit frequency per unit solid angle crossing a surface perpendicular to the beam: i.e.,

$$I_\nu = \frac{h\nu\, \delta N}{\mathcal{A}\, \delta t\, \Delta\nu\, \Delta\Omega}.$$

Direct comparison reveals $\mathcal{N} = h^{-4}(I_\nu/\nu^3)$.

Thus, conservation of $\mathcal{N}$ along a photon's world line implies conservation of I_ν/ν^3. This conservation law finds important applications in cosmology (e.g., Box 29.2 and Ex. 29.5) and in the gravitational lens effect (Refsdal 1964); see also exercises 22.15–22.17.

Exercise 22.18. STRESS-ENERGY TENSOR

(a) Show that the stress-energy tensor for a swarm of identical particles at an event $\mathcal{P}_0$ can be written as an integral over the mass hyperboloid of the momentum space at $\mathcal{P}_0$:

$$\boldsymbol{T} = \int (\mathcal{N} \boldsymbol{p} \otimes \boldsymbol{p})(d\mathcal{V}_p/p^0), \tag{22.53}$$

$$\frac{d\mathcal{V}_p}{p^0} \equiv \frac{dp^x\, dp^y\, dp^z}{p^0} \text{ in a local Lorentz frame.} \tag{22.54}$$

(Notice from Box 22.5 that $d\mathcal{V}_p/p^0$ is a Lorentz-invariant volume element for any segment of the mass hyperboloid.)

(b) Verify that the Boltzmann equation, $d\mathcal{N}/d\lambda = 0$, implies $\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$ for any swarm of identical particles. [*Hint:* Calculate $\boldsymbol{\nabla}\cdot\boldsymbol{T}$ in a local Lorentz frame, using the above expression for $\boldsymbol{T}$, and using the geodesic equation in the form $Dp^\mu/d\lambda = 0$.]

Exercise 22.19. KINETIC THEORY FOR NONIDENTICAL PARTICLES

For a swarm of particles with a wide distribution of rest masses, define

$$\mathcal{N} = \frac{\Delta N}{\mathcal{V}_x \mathcal{V}_p\, \Delta m}, \tag{22.55}$$

where $\mathcal{V}_x$ and $\mathcal{V}_p$ are spatial and momentum 3-volumes, and ΔN is the number of particles in the region $\mathcal{V}_x\mathcal{V}_p$ with rest masses between $m - \Delta m/2$ and $m + \Delta m/2$. Show the following.

(a) $\mathcal{V}_x\mathcal{V}_p\,\Delta m$ is independent of Lorentz frame and independent of location on the world tube of a bundle of particles.

(b) $\mathcal{N}$ can be regarded as a function of location $\mathcal{P}$ in spacetime and 4-momentum $\boldsymbol{p}$ inside the future light cone of the tangent space at $\mathcal{P}$:

$$\mathcal{N} = \mathcal{N}(\mathcal{P}, \boldsymbol{p}). \tag{22.56}$$

(c) $\mathcal{N}$ satisfies the collisionless Boltzmann equation (kinetic equation)

$$\frac{d\mathcal{N}[\mathcal{P}(\lambda), \boldsymbol{p}(\lambda)]}{d\lambda} = 0 \qquad \text{along geodesic trajectory of any particle.} \tag{22.57}$$

(d) $\mathcal{N}$ can be rewritten in a local Lorentz frame as

$$\mathcal{N} = \frac{\Delta N}{[(p^0/m)\,\Delta x\,\Delta y\,\Delta z][\Delta p^0\,\Delta p^x\,\Delta p^y\,\Delta p^z]}. \tag{22.58}$$

(e) The stress-energy tensor at an event $\mathcal{P}$ can be written as an integral over the interior of the future light cone of momentum space

$$T^{\mu\nu} = \int (\mathcal{N} p^\mu p^\nu) m^{-1}\, dp^0\, dp^1\, dp^2\, dp^3 \tag{22.59}$$

in a local Lorentz frame (Track-1 notation for integral; see Box 5.3);

$$\boldsymbol{T} = \int (\mathcal{N}\boldsymbol{p}\otimes\boldsymbol{p}) m^{-1}\, {}^*1 \qquad \text{in frame-independent notation}$$

$$= \int (\mathcal{N}\boldsymbol{p}\otimes\boldsymbol{p}) m^{-1}\, \boldsymbol{d}p^0 \wedge \boldsymbol{d}p^1 \wedge \boldsymbol{d}p^2 \wedge \boldsymbol{d}p^3 \tag{22.59$'$}$$

in a local Lorentz frame (Track-2 notation; see Box 5.4).