

Elementary General Relativity

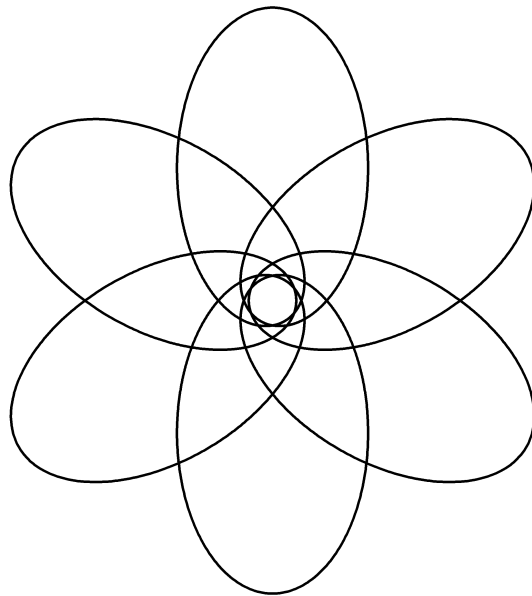
Version 3.35

Alan Macdonald

Luther College, Decorah, IA USA

<mailto:macdonal@luther.edu>

<http://faculty.luther.edu/~macdonal>



To Ellen

“The magic of this theory will hardly fail to impose itself on anybody who has truly understood it.”

Albert Einstein, 1915

“The foundation of general relativity appeared to me then [1915], and it still does, the greatest feat of human thinking about Nature, the most amazing combination of philosophical penetration, physical intuition, and mathematical skill.”

Max Born, 1955

“One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity.”

Josiah Willard Gibbs

“There is a widespread indifference to attempts to put accepted theories on better logical foundations and to clarify their experimental basis, an indifference occasionally amounting to hostility. I am concerned with the effects of our neglect of foundations on the education of scientists. It is plain that the clearer the teacher, the more transparent his logic, the fewer and more decisive the number of experiments to be examined in detail, the faster will the pupil learn and the surer and sounder will be his grasp of the subject.”

Sir Hermann Bondi

“Things should be made as simple as possible, but not simpler.”

Albert Einstein

Contents

Preface

1	Flat Spacetimes	
1.1	Spacetimes	11
1.2	The Inertial Frame Postulate	14
1.3	The Metric Postulate	18
1.4	The Geodesic Postulate	25
2	Curved Spacetimes	
2.1	History of Theories of Gravity	27
2.2	The Key to General Relativity	30
2.3	The Local Inertial Frame Postulate	34
2.4	The Metric Postulate	37
2.5	The Geodesic Postulate	40
2.6	The Field Equation	43
3	Spherically Symmetric Spacetimes	
3.1	Stellar Evolution	49
3.2	The Schwarzschild Metric	51
3.3	The Solar System Tests	56
3.4	Kerr Spacetimes	60
3.5	The Binary Pulsar	62
3.6	Black Holes	63
4	Cosmological Spacetimes	
4.1	Our Universe I	66
4.2	The Robertson-Walker Metric	69
4.3	The Expansion Redshift	71
4.4	Our Universe II	73
4.5	General Relativity Today	79
	Appendices	81
	Index	101

Preface

The purpose of this book is to provide, with a minimum of mathematical machinery and in the fewest possible pages, a clear and careful explanation of the physical principles and applications of classical general relativity. The prerequisites are single variable calculus, a few basic facts about partial derivatives and line integrals, a little matrix algebra, and some basic physics.

The book is for those seeking a conceptual understanding of the theory, not computational prowess. Despite its brevity and modest prerequisites, it is a serious introduction to the physics and mathematics of general relativity which demands careful study. The book can stand alone as an introduction to general relativity or it can be used as an adjunct to standard texts.

Chapter 1 is a self-contained introduction to those parts of special relativity we require for general relativity. We take a nonstandard approach to the metric, analogous to the standard approach to the metric in Euclidean geometry. In geometry, distance is first understood *geometrically*, independently of any coordinate system. If coordinates are introduced, then distances can be expressed in terms of coordinate differences: $\Delta s^2 = \Delta x^2 + \Delta y^2$. The formula is important, but the geometric meaning of the distance is fundamental.

Analogously, we define the spacetime interval of special relativity *physically*, independently of any coordinate system. If inertial frame coordinates are introduced, then the interval can be expressed in terms of coordinate differences: $\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$. The formula is important, but the physical meaning of the interval is fundamental. I believe that this approach to the metric provides easier access to and deeper understanding of special relativity, and facilitates the transition to general relativity.

Chapter 2 introduces the physical principles on which general relativity is based. The basic concepts of Riemannian geometry are developed in order to express these principles mathematically as postulates. The purpose of the postulates is not to achieve complete rigor – which is neither desirable nor possible in a book at this level – but to state clearly the physical principles, and to exhibit clearly the relationship to special relativity and the analogy with surfaces. The postulates are in one-to-one correspondence with the fundamental concepts of Riemannian geometry: manifold, metric, geodesic, and curvature. Concentrating on the physical meaning of the metric greatly simplifies the development of general relativity. In particular, tensors are not needed. There is, however, a brief introduction to tensors in an appendix. (Similarly, modern elementary differential geometry texts often develop the intrinsic geometry of curved surfaces by focusing on the geometric meaning of the metric. Tensors are not used.)

The first two chapters systematically exploit the mathematical analogy which led to general relativity: a curved spacetime is to a flat spacetime as a curved surface is to a flat surface. Before introducing a spacetime concept, its analog

for surfaces is presented. This is not a new idea, but it is used here more systematically than elsewhere. For example, when the metric ds of general relativity is introduced, the reader has already seen a metric in three other contexts.

Chapter 3 solves the field equation for a spherically symmetric spacetime to obtain the Schwarzschild metric. The geodesic equations are then solved and applied to the classical solar system tests of general relativity. There is a section on the Kerr metric, including gravitomagnetism and the Gravity Probe B experiment. The chapter closes with sections on the binary pulsar and black holes. In this chapter, as elsewhere, I have tried to provide the cleanest possible calculations.

Chapter 4 applies general relativity to cosmology. We obtain the Robertson-Walker metric in an elementary manner without using the field equation. We then solve the field equation with a nonzero cosmological constant for a flat Robertson-Walker spacetime. WMAP data allow us to specify all parameters in the solution, giving the new “standard model” of the universe with dark matter and dark energy.

There have been many spectacular astronomical discoveries and observations since 1960 which are relevant to general relativity. We describe them at appropriate places in the book.

Some 50 exercises are scattered throughout. They often serve as examples of concepts introduced in the text. If they are not done, they should be read.

Some tedious (but always straightforward) calculations have been omitted. They are best carried out with a computer algebra system. Some material has been placed in about 20 pages of appendices to keep the main line of development visible. The appendices occasionally require more background of the reader than the text. They may be omitted without loss of anything essential. Appendix 1 gives the values of various physical constants. Appendix 2 contains several approximation formulas used in the text.

Chapter 1

Flat Spacetimes

1.1 Spacetimes

The general theory of relativity is our best theory of space, time, and gravity. It is commonly felt to be the most beautiful of all physical theories. Albert Einstein created the theory during the decade following the publication, in 1905, of his special theory of relativity. The special theory is a theory of space and time which applies when gravity is insignificant. The general theory generalizes the special theory to include gravity.

In geometry the fundamental entities are points. A point is a specific place. In relativity the fundamental entities are *events*. An event is a specific time and place. For example, the collision of two particles is an event. A concert is an event (idealizing it to a single time and place). To attend the concert, you must be at the time *and* the place of the event.

A flat or curved surface is a set of points. (We shall prefer the term “flat surface” to “plane”.) Similarly, a *spacetime* is a set of events. For example, we might consider the events in a specific room between two specific times. A *flat* spacetime is one without significant gravity. Special relativity describes flat spacetimes. A *curved* spacetime is one with significant gravity. General relativity describes curved spacetimes.

There is nothing mysterious about the words “flat” or “curved” attached to a set of events. They are chosen because of a remarkable analogy, already hinted at, concerning the mathematical description of a spacetime: a curved spacetime is to a flat spacetime as a curved surface is to a flat surface. This analogy will be a major theme of this book; we shall use the mathematics of flat and curved surfaces to guide our understanding of the mathematics of flat and curved spacetimes.

We shall explore spacetimes with clocks to measure time and rods (rulers) to measure space, i.e., distance. However, clocks and rods do not in fact live up to all that we usually expect of them. In this section we shall see what we expect of them in relativity.

Clocks. A curve is a continuous succession of points in a surface. Similarly, a *worldline* is a continuous succession of events in a spacetime. A moving particle or a pulse of light emitted in a single direction is present at a continuous succession of events, its worldline. Even if a particle is at rest, time passes, and the particle has a worldline.

The length of a curve between two given points depends on the curve. Similarly, the time between two given events measured by a clock moving between the events depends on the clock's worldline! J. C. Hafele and R. Keating provided the most direct verification of this in 1971. They brought two atomic clocks together, then placed them in separate airplanes which circled the Earth in opposite directions, and then brought the clocks together again. Thus the clocks took different worldlines between the event of their separation and the event of their reunion. They measured different times between the two events. The difference was small, about 10^{-7} sec, but was well within the ability of the clocks to measure. There is no doubt that the effect is real.

Relativity predicts the measured difference. Exercise 1.10 shows that special relativity predicts a difference between the clocks. Exercise 2.1 shows that general relativity predicts a further difference. Exercise 3.3 shows that general relativity predicts the observed difference. Relativity predicts large differences between clocks whose relative velocity is close to the speed of light.

The best answer to the question “How can the clocks in the experiment *possibly* disagree?” is the question “Why *should* they agree?” After all, the clocks are not connected. According to everyday ideas they should agree because there is a universal time, common to all observers. It is the duty of clocks to report this time. The concept of a universal time was abstracted from experience with small speeds (compared to that of light) and clocks of ordinary accuracy, where it is nearly valid. The concept permeates our daily lives; there are clocks everywhere telling us *the* time. However, the Hafele-Keating experiment shows that there is no universal time. Time, like distance, is route dependent.

Since clocks on different worldlines between two events measure different times between the events, we cannot speak of *the* time between two events. However, the *relative* rates of processes – the ticking of a clock, the frequency of a tuning fork, the aging of an organism, etc. – are the same along all worldlines. (Unless some adverse physical condition affects a rate.) Twins traveling in the two airplanes of the Hafele-Keating experiment would each age according to the clock in their airplane. They would thus be of slightly different ages when reunited.

Rods. Consider astronauts in interstellar space, where gravity is insignificant. If their rocket is not firing and their ship is not spinning, then they will feel no forces acting on them and they can float freely in their cabin. If their spaceship is accelerating, then they will feel a force pushing them back against their seat. If the ship turns to the left, then they will feel a force to the right. If the ship is spinning, they will feel a force outward from the axis of spin. Call these forces *inertial forces*.

Accelerometers measure inertial forces. Fig. 1.1 shows a simple accelerometer consisting of a weight held at the center of a frame by identical springs. Inertial forces cause the weight to move from the center. An *inertial object* is one which experiences no inertial forces. If an object is inertial, then any object moving at a constant velocity with respect to it is also inertial and any object accelerating with respect to it is not inertial.

In special relativity we make an assumption which allows us to speak of *the* distance between two inertial objects at rest with respect to each other: Inertial rigid rods side by side and at rest with respect to two inertial objects measure the same distance between the objects. More precisely, we assume that any difference is due to some adverse physical cause (e.g., thermal expansion) to which an “ideal” rigid rod would not be subject. In particular, the history of a rigid rod does not affect its length. Noninertial rods are difficult to deal with in relativity, and we shall not consider them.

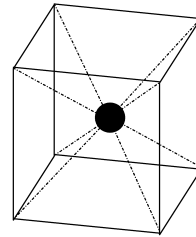


Fig. 1.1: An accelerometer. The weight is held at the center by springs. Acceleration causes the weight to move from the center.

In the next three sections we give three postulates for special relativity. The *inertial frame postulate* asserts that certain natural coordinate systems, called inertial frames, exist for a flat spacetime. The *metric postulate* asserts a universal light speed and a slowing of clocks moving in inertial frames. The *geodesic postulate* asserts that inertial particles and light move in a straight line at constant speed in inertial frames.

We shall use the analogy mentioned above to help us understand the postulates. Imagine two dimensional beings living in a flat surface. These *surface dwellers* can no more imagine leaving their two spatial dimensions than we can imagine leaving our three spatial dimensions. Before introducing a postulate for a flat spacetime, we introduce the analogous postulate formulated by surface dwellers for a flat surface. The postulates for a flat spacetime use a time dimension, but those for a flat surface do not.

1.2 The Inertial Frame Postulate

Surface dwellers find it useful to label the points of their flat surface with coordinates. They construct, using identical rigid rods, a square grid and assign rectangular coordinates (x, y) to the nodes of the grid in the usual way. See Fig. 1.2. They specify a point by using the coordinates of the node nearest the point. If more accurate coordinates are required, they build a finer grid. Surface dwellers call the coordinate system a *planar frame*. They postulate:

The Planar Frame Postulate for a Flat Surface

A planar frame can be constructed with any point P as origin and with any orientation.

Similarly, it is useful to label the events in a flat spacetime with coordinates (t, x, y, z) . The coordinates specify *when* and *where* the event occurs, i.e., they completely specify the event. We now describe how to attach coordinates to events. The procedure is idealized, but it gives a clear physical meaning to the coordinates.

To specify *where* an event occurs, construct, using identical rigid rods, an inertial cubical lattice. See Fig. 1.3. Assign rectangular coordinates (x, y, z) to the nodes of the lattice in the usual way. Specify where an event occurs by using the coordinates of the node nearest the event.

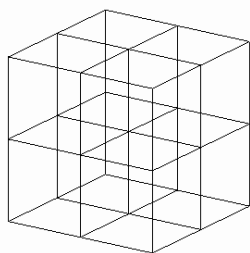


Fig. 1.3: An inertial lattice.

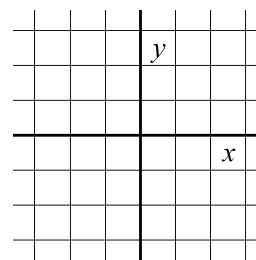


Fig. 1.2: A planar frame.

To specify *when* an event occurs, place a clock at each node of the lattice. Then the times of events at a given node can be specified by reading the clock at that node. But to compare meaningfully the times of events at different nodes, the clocks must be in some sense synchronized. As we shall see soon, this is not a trivial matter. (Remember, there is no universal time.) For now, assume that the clocks have been synchronized. Then specify when an event occurs by using the time, t , on the clock at the node nearest the event. And measure the coordinate time difference Δt between two events using the synchronized clocks at the nodes where the events occur. Note that this requires *two* clocks.

The four dimensional coordinate system obtained in this way from an inertial cubical lattice with synchronized clocks is called an *inertial frame*. The event $(t, x, y, z) = (0, 0, 0, 0)$ is the *origin* of the inertial frame. We postulate:

The Inertial Frame Postulate for a Flat Spacetime

An inertial frame can be constructed with any event E as origin, with any orientation, and with any inertial object at E at rest in it.

If we suppress one or two of the spatial coordinates of an inertial frame, then we can draw a *spacetime diagram* and depict worldlines. For example, Fig. 1.4 shows the worldlines of two particles. One is at rest on the x -axis and one moves away from $x = 0$ and then returns more slowly.

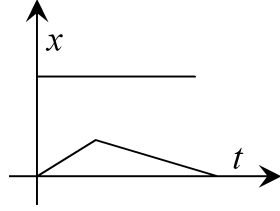


Fig. 1.4: Two worldlines.

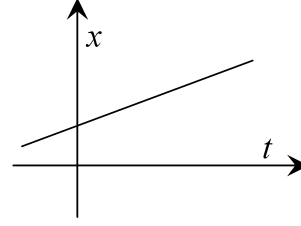


Fig. 1.5: Worldline of a particle moving with constant speed.

Exercise 1.1. Show that the worldline of an object moving along the x -axis at constant speed v is a straight line with slope v . See Fig. 1.5.

Exercise 1.2. Describe the worldline of an object moving in a circle in the $z = 0$ plane at constant speed v .

Synchronization. We return to the matter of synchronizing the clocks in the lattice. What does it mean to say that separated clocks are synchronized? Einstein realized that the answer to this question is not given to us by Nature; rather, it must be answered with a definition.

Exercise 1.3. Why not simply bring the clocks together, synchronize them, move them to the nodes of the lattice, and call them synchronized?

We might try the following definition. Send a signal from a node P of the lattice at time t_P according to the clock at P . Let it arrive at a node Q of the lattice at time t_Q according to the clock at Q . Let the distance between the nodes be D and the speed of the signal be v . Say that the clocks are synchronized if

$$t_Q = t_P + D/v. \quad (1.1)$$

Intuitively, the term D/v compensates for the time it takes the signal to get to Q . This definition is flawed because it contains a logical circle: v is *defined* by a rearrangement of Eq. (1.1): $v = D/(t_Q - t_P)$. Synchronized clocks cannot be defined using v because synchronized clocks are needed to define v .

We adopt the following definition, essentially due to Einstein. Emit a pulse of light from a node P at time t_P according to the clock at P . Let it arrive at a node Q at time t_Q according to the clock at Q . Similarly, emit a pulse of light from Q at time t'_Q and let it arrive at P at t'_P . The clocks are *synchronized* if

$$t_Q - t_P = t'_P - t'_Q, \quad (1.2)$$

i.e., if the times in the two directions are equal.

Reformulating the definition makes it more transparent. If the pulse from Q to P is the reflection of the pulse from P to Q , then $t'_Q = t_Q$ in Eq. (1.2). Let $2T$ be the round trip time: $2T = t'_P - t_P$. Substitute this in Eq. (1.2):

$$t_Q = t_P + T; \quad (1.3)$$

the clocks are synchronized if the pulse arrives at Q in half the time it takes for the round trip.

Exercise 1.4. Explain why Eq. (1.2) is a satisfactory definition but Eq. (1.1) is not.

There is a tacit assumption in the definition of synchronized clocks that the two sides of Eq. (1.2) do not depend on the times that the pulses are sent:

Emit pulses of light from a node R at times t_R and t'_R according to a clock at R . Let them arrive at a node S at times t_S and t'_S according to a clock at S . Then

$$t'_S - t'_R = t_S - t_R. \quad (1.4)$$

With this assumption we can be sure that synchronized clocks will remain synchronized.

Exercise 1.5. Show that with the assumption Eq. (1.4), T in Eq. (1.3) is independent of the time the pulse is sent.

A rearrangement of Eq. (1.4) gives

$$\Delta s_o = \Delta s_e, \quad (1.5)$$

where $\Delta s_o = t'_S - t_S$ is the time between the observation of the pulses at S and $\Delta s_e = t'_R - t_R$ is the time between the emission of the pulses at R . (We use Δs rather than Δt to conform to notation used later in more general situations.) If a clock at R emits pulses of light at regular intervals to S , then Eq. (1.5) states that an observer at S sees (actually *sees*) the clock at R going at the same rate as his clock. Of course, the observer at S will see all physical processes at R proceed at the same rate they do at S .

Redshifts. We will encounter situations in which $\Delta s_o \neq \Delta s_e$. Define the *redshift*

$$z = \frac{\Delta s_o}{\Delta s_e} - 1. \quad (1.6)$$

Equations (1.4) and (1.5) correspond to $z = 0$. If, for example, $z = 1$ ($\Delta s_o/\Delta s_e = 2$), then the observer at S would see clocks at R , and all other physical processes at R , proceed at half the rate they do at S .

If the two “pulses” of light in Eq. (1.6) are successive wavecrests of light emitted at frequency $f_e = (\Delta s_e)^{-1}$ and observed at frequency $f_o = (\Delta s_o)^{-1}$, then Eq. (1.6) can be written

$$z = \frac{f_e}{f_o} - 1. \quad (1.7)$$

In Exercise 1.6 we shall see that Eq. (1.5) is violated, i.e., $z \neq 0$, if the emitter and observer are in relative motion in a flat spacetime. This is called a *Doppler redshift*. Later we shall see two other kinds of redshift: *gravitational redshifts* in Sec. 2.2 and *expansion redshifts* in Sec. 4.1. The three types of redshifts have different physical origins and so must be carefully distinguished.

Synchronization. The inertial frame postulate asserts in part that clocks in an inertial lattice can be synchronized according to the definition Eq. (1.2), or, in P. W. Bridgeman's descriptive phrase, we can "spread time over space". We now prove this with the aid of an auxiliary assumption. The reader may skip the proof and turn to the next section without loss of continuity.

Let $2T$ be the time, as measured by a clock at the origin O of the lattice, for light to travel from O to another node Q and return after being reflected at Q . Emit a pulse of light at O toward Q at time t_O according to the clock at O . When the pulse arrives at Q set the clock there to $t_Q = t_O + T$. According to Eq. (1.3) the clocks at O and Q are now synchronized.

Synchronize all clocks with the one at O in this way.

To show that the clocks at any two nodes P and Q are now synchronized with each other, we must make this assumption:

The time it takes light to traverse a triangle in the lattice is independent of the direction taken around the triangle.

See Fig. 1.6. In an experiment performed in 1963, W. M. Macek and D. T. M. Davis, Jr. verified the assumption for a square to one part in 10^{12} . See Appendix 3.

Reflect a pulse of light around the triangle OPQ . Let the pulse be at O, P, Q, O at times t_O, t_P, t_Q, t_R according to the clock at that node. Similarly, let the times for a pulse sent around in the other direction be t'_O, t'_P, t'_Q, t'_R . See Fig. 1.6. We have the algebraic identities

$$\begin{aligned} t_R - t_O &= (t_R - t_P) + (t_P - t_Q) + (t_Q - t_O) \\ t'_R - t'_O &= (t'_R - t'_P) + (t'_P - t'_Q) + (t'_Q - t'_O). \end{aligned} \quad (1.8)$$

According to our assumption, the left sides of the two equations are equal. Also, since the clock at O is synchronized with those at P and Q ,

$$t_P - t_O = t'_P - t'_O \quad \text{and} \quad t_R - t_Q = t'_R - t'_Q.$$

Thus, subtracting the equations Eq. (1.8) shows that the clocks at P and Q are synchronized:

$$t_Q - t_P = t'_Q - t'_P.$$

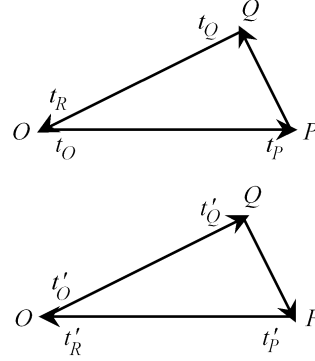


Fig. 1.6: Light traversing a triangle in opposite directions.

1.3 The Metric Postulate

Let P and Q be points in a flat surface. Different curves between the points have different lengths. But surface dwellers single out for special study the length Δs of the straight line between P and Q . They call Δs the *proper distance* between the points.

The proper distance Δs between two points is defined *geometrically*, independently of any planar frame. But there is a simple formula for Δs in terms of the coordinate differences between the points in a planar frame:

The Metric Postulate for a Flat Surface

Let Δs be the proper distance between points P and Q . Let P and Q have coordinate differences $(\Delta x, \Delta y)$ in a planar frame. Then

$$\Delta s^2 = \Delta x^2 + \Delta y^2. \quad (1.9)$$

The coordinate differences Δx and Δy between P and Q are different in different planar frames. See Fig. 1.7. However, the particular combination of the differences in Eq. (1.9) will always produce Δs . Neither Δx nor Δy has a geometric significance independent of the particular planar frame chosen. The two of them together do: they determine Δs , which has a direct geometric significance, independent of any coordinate system.

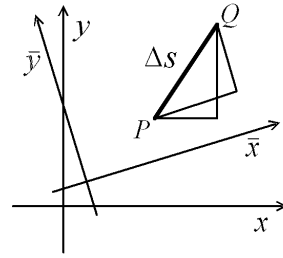


Fig. 1.7: $\Delta s^2 = \Delta x^2 + \Delta y^2$ in both planar frames.

Let E and F be events in a flat spacetime. There is a distance-like quantity Δs between them. It is called the (*spacetime*) *interval* between the events. The definition of Δs in a flat spacetime is more complicated than in a flat surface, as there are three ways in which events can be, we say, *separated*:

- If E and F can be on the worldline of a pulse of light, they are *lightlike separated*. Then define $\Delta s = 0$.
- If E and F can be on the worldline of an inertial clock, they are *timelike separated*. Then define Δs to be the time between the events measured by the clock. This is the *proper time* between the events. Other clocks moving between the events will measure different times. But we single out for special study the proper time Δs .
- If E and F can be simultaneously at the ends of an inertial rigid rod – simultaneously in the sense that light flashes emitted at E and F reach the center of the rod simultaneously, or equivalently, that E and F are simultaneous in the rest frame of the rod – they are *spacelike separated*. Then define $|\Delta s|$ to be the length the rod. (The reason for the absolute value will become clear later.) This is the *proper distance* between the events.

The spacetime interval Δs between two events is defined *physically*, independently of any inertial frame. But there is a simple formula for Δs in terms of the coordinate differences between the events in an inertial frame:

The Metric Postulate for a Flat Spacetime

Let Δs be the interval between events E and F . Let the events have coordinate differences $(\Delta t, \Delta x, \Delta y, \Delta z)$ in an inertial frame. Then

$$\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2. \quad (1.10)$$

The coordinate differences between E and F , including the time coordinate difference, are different in different inertial frames. For example, suppose that an inertial clock measures a proper time Δs between two events. In an inertial frame in which the clock is at rest, $\Delta t = \Delta s$ and $\Delta x = \Delta y = \Delta z = 0$. This will not be the case in an inertial frame in which the clock is moving. However, the particular combination of the differences in Eq. (1.10) will always produce Δs . No one of the coordinate differences has a physical significance independent of the particular inertial frame chosen. The four of them together do: they determine Δs , which has a direct physical significance, independent of any inertial frame.

This shows that the joining of space and time into spacetime is not an artificial technical trick. Rather, in the words of Hermann Minkowski, who introduced the spacetime concept in 1908, “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.”

Physical Meaning. We now describe the physical meaning of the metric postulate for lightlike, timelike, and spacelike separated events. We do not need the y - and z -coordinates for the discussion, and so we use Eq. (1.10) in the form

$$\Delta s^2 = \Delta t^2 - \Delta x^2. \quad (1.11)$$

Lightlike separated events. By definition, a pulse of light can move between lightlike separated events, and $\Delta s = 0$ for the events. From Eq. (1.11) the speed of the pulse is $|\Delta x|/\Delta t = 1$. *The metric postulate asserts that the speed c of light has always the same value $c = 1$ in all inertial frames.* The important thing is that the speed is the same in all inertial frames; the actual value $c = 1$ is then a convention: Choose the distance light travels in one second – about 3×10^{10} cm – as the unit of distance. Call this one (*light*) *second* of distance. (You are probably familiar with a similar unit of distance – the light year.) Then $1 \text{ cm} = 3.3 \times 10^{-11} \text{ sec}$. With this convention $c = 1$, and all other speeds are expressed as a fraction of the speed of light. Ordinarily the fractions are very small.

Timelike separated events. By definition, an inertial clock can move between timelike separated events, and Δs is the (proper) time the clock measures between the events. The speed of the clock is $v = |\Delta x|/\Delta t$. Then from Eq. (1.11), the proper time is

$$\Delta s = (\Delta t^2 - \Delta x^2)^{\frac{1}{2}} = [1 - (\Delta x/\Delta t)^2]^{\frac{1}{2}} \Delta t = (1 - v^2)^{\frac{1}{2}} \Delta t. \quad (1.12)$$

The metric postulate asserts that the proper time between two events is less than the time determined by the synchronized clocks of an inertial frame: $\Delta s < \Delta t$. Informally, “moving clocks run slowly”. This is called *time dilation*. According to Eq. (1.12) the time dilation factor is $(1 - v^2)^{\frac{1}{2}}$. For normal speeds, v is very small, v^2 is even smaller, and so from Eq. (1.12), $\Delta s \approx \Delta t$, as expected. But as $v \rightarrow 1$, $\Delta t/\Delta s = (1 - v^2)^{-\frac{1}{2}} \rightarrow \infty$. Fig. 1.8 shows the graph of $\Delta t/\Delta s$ vs. v .

Exercise 1.6. Investigate the Doppler redshift. Let a source of light pulses move with velocity v directly away from an observer at rest in an inertial frame. Let Δt_e be the time between the emission of pulses, and Δt_o be the time between the reception of pulses by the observer.

- Show that $\Delta t_o = \Delta t_e + v\Delta t_e/c$.
- Ignore time dilation in Eq. (1.6) by setting $\Delta s = \Delta t$. Show that $z = v/c$ in this approximation.
- Show that the exact formula is $z = [(1 + v)/(1 - v)]^{\frac{1}{2}} - 1$ (with $c = 1$). Use the result of part a.

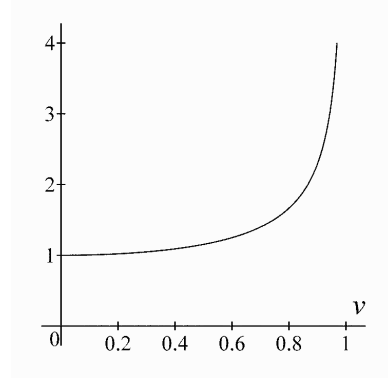


Fig. 1.8: $\Delta t/\Delta s = (1 - v^2)^{-\frac{1}{2}}$.

Spacelike separated events. By definition, the ends of an inertial rigid rod can be simultaneously present at the events, and $|\Delta s|$ is the length of the rod. Appendix 4 shows that the speed of the rod in I is $v = \Delta t/|\Delta x|$. (That is not a typo.) A calculation similar to Eq. (1.12) gives

$$|\Delta s| = (1 - v^2)^{\frac{1}{2}} |\Delta x|. \quad (1.13)$$

The metric postulate asserts that the proper distance between spacelike separated events is less than an inertial frame distance: $|\Delta s| < |\Delta x|$. (This is not length contraction, which we discuss in Appendix 5.)

Connections. We have just seen that the physical meaning of the metric postulate is different for lightlike, timelike, and spacelike separated events. However, the meanings are connected: the physical meaning for lightlike separated events (a universal light speed) implies the physical meanings for timelike and spacelike separated events. We now prove this for the timelike case. The proof is quite instructive. The spacelike case is less so; it is relegated to Appendix 4.

By definition, an inertial clock C can move between time-like separated events E and F , and Δs is the time C measures between the events. Let C carry a rod R perpendicular to its direction of motion. Let R have a mirror M on its end. At E a light pulse is sent along R from C . The length of R is chosen so that the pulse is reflected by M back to F . Fig. 1.9 shows the path of the light in I , together with C , R , and M as the light reflects off M .

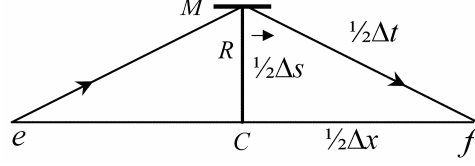


Fig. 1.9: $\Delta s^2 = \Delta t^2 - \Delta x^2$ for timelike separated events. e and f are the points at which events E and F occur.

Refer to the rightmost triangle in Fig. 1.9. In I , the distance between E and F is Δx . This gives the labeling of the base of the triangle. In I , the light takes the time Δt from E to M to F . Since $c = 1$ in I , the light travels a distance Δt in I . This gives the labeling of the hypotenuse. C is at rest in some inertial frame I' . In I' , the light travels the length of the rod twice in the proper time Δs between E and F measured by C . Since $c = 1$ in I' , the length of the rod is $\frac{1}{2}\Delta s$ in I' . This gives the labeling of the altitude of the triangle. (There is a tacit assumption here that the length of R is the same in I and I' . Appendix 5 discusses this.) Applying the Pythagorean theorem to the triangle shows that Eq. (1.11) is satisfied for timelike separated events.

In short, since the light travels farther in I than in I' (the hypotenuse twice vs. the altitude twice) and the speed $c = 1$ is the same in I and I' , the time (= distance/speed) between E and F is longer in I than for C . This shows, in a most graphic way, that accepting a universal light speed forces us to abandon a universal time.

The argument shows how it is possible for a single pulse of light to have the same speed in inertial frames moving with respect to each other: the speed (= distance/time) can be the same because the distance *and* the time are different in the two frames.

Exercise 1.7. Criticize the following argument. We have just seen that the time between two events is greater in I than in I' . But exactly the same argument carried out in I' will show that the time between the events is greater in I' than in I . This is a contradiction.

Local Forms. The metric postulate for a planar frame, Eq. (1.9), gives only the distance along a *straight line* between two points. The differential version of Eq. (1.9) gives the distance ds between *neighboring* points along *any* curve:

The Metric Postulate for a Flat Surface, Local Form

Let P and Q be neighboring points. Let ds be the distance between them. Let the points have coordinate differences (dx, dy) in a planar frame. Then

$$ds^2 = dx^2 + dy^2.$$

Thus, if a curve is parameterized $(x(p), y(p))$, $a \leq p \leq b$, then

$$ds^2 = \left[\left(\frac{dx}{dp} \right)^2 + \left(\frac{dy}{dp} \right)^2 \right] dp^2,$$

and the length of the curve is

$$s = \int_{p=a}^b ds = \int_{p=a}^b \left[\left(\frac{dx}{dp} \right)^2 + \left(\frac{dy}{dp} \right)^2 \right]^{\frac{1}{2}} dp.$$

Of course, different curves between two points can have different lengths.

Exercise 1.8. Calculate the circumference of the circle $x = r \cos \theta$, $y = r \sin \theta$, $0 \leq \theta \leq 2\pi$.

The metric postulate for an inertial frame Eq. (1.10) is concerned only with times measured by *inertial* clocks. The differential version of Eq. (1.10) gives the time ds measured by *any* clock between *neighboring* events on its worldline:

The Metric Postulate for a Flat Spacetime, Local Form

Let E and F be neighboring events. If E and F are lightlike separated, let $ds = 0$. If the events are timelike separated, let ds be the time between them as measured by any (inertial or noninertial) clock. Let the events have coordinate differences (dt, dx, dy, dz) in an inertial frame. Then

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2. \quad (1.14)$$

From Eq. (1.14), if the worldline of a clock is parameterized

$$(t(p), x(p), y(p), z(p)), \quad a \leq p \leq b,$$

then the time s to traverse the worldline, as measured by the clock, is

$$s = \int_{p=a}^b ds = \int_{p=a}^b \left[\left(\frac{dt}{dp} \right)^2 - \left(\frac{dx}{dp} \right)^2 - \left(\frac{dy}{dp} \right)^2 - \left(\frac{dz}{dp} \right)^2 \right]^{\frac{1}{2}} dp.$$

In general, clocks on different worldlines between two events will measure different times between the events.

Exercise 1.9. Let a clock move between two events with a time difference Δt . Let v be the small constant speed of the clock. Show that $\Delta t - \Delta s \approx \frac{1}{2}v^2 \Delta t$.

Exercise 1.10. Consider a simplified Hafele-Keating experiment. One clock remains on the ground and the other circles the equator in an airplane to the west – opposite to the Earth’s rotation. Assume that the Earth is spinning on its axis at one revolution per 24 hours in an inertial frame. (Thus the clock on the ground is not at rest.) Notation: Δt is the duration of the trip in the

inertial frame, v_r is the velocity of the clock remaining on the ground, and Δs_r is the time it measures for the trip. The quantities v_a and Δs_a are defined similarly for the airplane.

Use Exercise 1.9 for each clock to show that the difference between the clocks due to time dilation is $\Delta s_a - \Delta s_r = \frac{1}{2}(v_a^2 - v_r^2)\Delta t$. Suppose that $\Delta t = 40$ hours and the speed of the airplane with respect to the ground is 1000 km/hr. Substitute values to obtain $\Delta s_a - \Delta s_r = 1.4 \times 10^{-7}$ s.

Experimental Evidence. Because general relativistic effects play a part in the Hafele-Keating experiment (see Exercise 2.1), and because the uncertainty of the experiment is large ($\pm 10\%$), this experiment is not a precision test of time dilation for clocks. Much better evidence comes from observations of subatomic particles called *muons*. When at rest the average lifetime of a muon is 3×10^{-6} sec. According to the differential version of Eq. (1.12), if the muon is moving in a circle with constant speed v , then its average life, as measured in the laboratory, should be larger by a factor $(1 - v^2)^{-\frac{1}{2}}$. An experiment performed in 1977 showed this within an experimental error of .2%. In the experiment $v = .9994$, giving a time dilation factor $\Delta t/\Delta s = (1 - v^2)^{-\frac{1}{2}} = 29$! The circular motion had an acceleration of 10^{21} cm/sec², and so this is a test of the local form Eq. (1.14) of the metric postulate as well as the original form Eq. (1.10).

There is excellent evidence for a universal light speed. First of all, realize that if clocks at P and Q are synchronized according to the definition Eq. (1.2), then the speed of light from P to Q is equal to the speed from Q to P . We emphasize that with our definition of synchronized clocks this equality is a matter of definition which can be neither confirmed nor refuted by experiment.

The speed c of light can be measured by sending a pulse of light from a point P to a mirror at a point Q at distance D and measuring the elapsed time $2T$ for it to return. Then $c = 2D/2T$; c is a two way speed, measured with a single clock. Equation 1.5 shows that this two way speed is equal to the one way speed from P to Q . Thus the one way speed of light can be measured by measuring the two way speed.

In a famous experiment performed in 1887, A. A. Michelson and E. W. Morley compared the two way speed of light in perpendicular directions from a given point. Their experiment has been repeated many times, most accurately by G. Joos in 1930, who found that any difference in the two way speeds is less than six parts in 10^{12} . The Michelson-Morley experiment is described in Appendix 6. A modern version of the experiment using lasers was performed in 1979 by A. Brillit and J. L. Hall. They found that any difference in the two way speed of light in perpendicular directions is less than four parts in 10^{15} . See Appendix 6.

Another experiment, performed by R. J. Kennedy and E. M. Thorndike in 1932, found the two way speed of light to be the same, within six parts in 10^9 , on days six months apart, during which time the Earth moved to the opposite side of its orbit. See Appendix 6. Inertial frames in which the Earth is at rest on days six months apart move with a relative speed of 60 km/sec (twice the Earth's orbital speed). A more recent experiment by D. Hils and Hall improved

the result by over two orders of magnitude.

These experiments provide good evidence that the two way speed of light is the same in different directions, places, and inertial frames and at different times. They thus provide strong motivation for our definition of synchronized clocks: If the two way speed of light has always the same value, what could be more natural than to define synchronized clocks by requiring that the one way speed have this value?

In all of the above experiments, the source of the light is at rest in the inertial frame in which the light speed is measured. If light were like baseballs, then the speed of a moving source would be imparted to the speed of light it emits. Strong evidence that this is not so comes from observations of certain neutron stars which are members of a binary star system and which emit X-ray pulses at regular intervals. These systems are described in Sec. 3.1. If the speed of the star were imparted to the speed of the X-rays, then various strange effects would be observed. For example, X-rays emitted when the neutron star is moving toward the Earth could catch up with those emitted earlier when it was moving away from the Earth, and the neutron star could be seen coming and going at the same time! See Fig. 1.10.

This does not happen; an analysis of the arrival times of the pulses at Earth made in 1977 by K. Brecher shows that no more than two parts in 10^9 of the speed of the source is added to the speed of the X-rays. (It is not possible to “see” the neutron star in orbit around its companion directly. The speed of the neutron star toward or away from the Earth can be determined from the Doppler redshift of the time between pulses. See Exercise 1.6.)

Finally, recall from above that the universal light speed statement of the metric postulate implies the statements about timelike and spacelike separated events. Thus the evidence for a universal light speed is also evidence for the other two statements.

The universal nature of the speed of light makes possible the modern definition of the unit of length: “The meter is the length of the path traveled by light during the time interval of $1/299,792,458$ of a second.” Thus, *by definition*, the speed of light is $299,792,458$ m/sec.

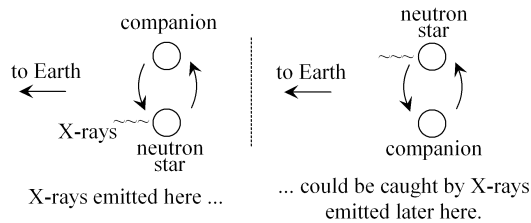


Fig. 1.10: The speed of light is independent of the speed of its source.

1.4 The Geodesic Postulate

We will find it convenient to use superscripts to distinguish coordinates. Thus we use (x^1, x^2) instead of (x, y) for planar frame coordinates.

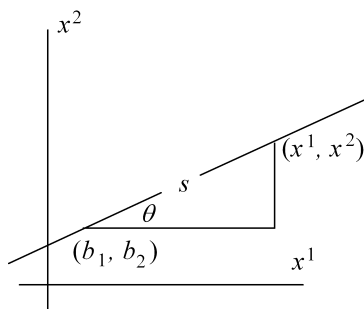


Fig. 1.11: A geodesic in a planar frame.

The line in Fig. 1.11 can be parameterized by the (proper) distance s from (b_1, b_2) to (x^1, x^2) : $x^1(s) = (\cos \theta)s + b_1$, $x^2(s) = (\sin \theta)s + b_2$. Differentiate twice with respect to s to obtain

The Geodesic Postulate for a Flat Surface

Parameterize a straight line with arclength s . Then in every planar frame

$$\ddot{x}^i = 0, \quad i = 1, 2. \quad (1.15)$$

The straight lines are called *geodesics*.

Not all parameterizations of a straight line satisfy the geodesic differential equations Eq. (1.15). Example: $x^i(p) = a_i p^3 + b_i$.

We will find it convenient to use (x^0, x^1, x^2, x^3) instead of (t, x, y, z) for inertial frame coordinates. Our third postulate for special relativity says that inertial particles and light pulses move in a straight line at constant speed in an inertial frame, i.e., their equations of motion are

$$x^i = a_i x^0 + b_i, \quad i = 1, 2, 3. \quad (1.16)$$

(Differentiate to give $dx^i/dx^0 = a_i$; the velocity is constant.) For inertial particles this is called Newton's first law.

Set $x^0 = p$, a parameter; $a_0 = 1$; and $b_0 = 0$, and find that worldlines of inertial particles and light can be parameterized

$$x^i(p) = a_i p + b_i, \quad i = 0, 1, 2, 3 \quad (1.17)$$

in an inertial frame. Eq. (1.17), unlike Eq. (1.16), is symmetric in all four coordinates of the inertial frame. Also, Eq. (1.17) shows that the worldline is a straight line in the spacetime. Thus “straight in spacetime” includes both “straight in space” and “straight in time” (constant speed). See Exercise 1.1. The worldlines are called *geodesics*.

Exercise 1.11. In Eq. (1.17) the parameter $p = x^0$. Show that the worldline of an inertial particle can also be parameterized with s , the proper time along the worldline.

The Geodesic Postulate for a Flat Spacetime

Worldlines of inertial particles and pulses of light can be parameterized with a parameter p so that in every inertial frame

$$\ddot{x}^i = 0, \quad i = 0, 1, 2, 3. \quad (1.18)$$

For inertial particles we may take $p = s$.

The geodesic postulate is a *mathematical* expression of our *physical* assertion that in an inertial frame inertial particles and light move in a straight line at constant speed.

Exercise 1.12. Make as long a list as you can of analogous properties of flat surfaces and flat spacetimes.

Chapter 2

Curved Spacetimes

2.1 History of Theories of Gravity

Recall the analogy from Chapter 1: A curved spacetime is to a flat spacetime as a curved surface is to a flat surface. We explored the analogy between a flat surface and a flat spacetime in Chapter 1. In this chapter we generalize from flat surfaces and flat spacetimes (spacetimes without significant gravity) to curved surfaces and curved spacetimes (spacetimes with significant gravity). General relativity interprets gravity as a curvature of spacetime.

Before embarking on a study of gravity in general relativity let us review, very briefly, the history of theories of gravity. These theories played a central role in the rise of science. Theories of gravity have their roots in attempts of the ancients to predict the motion of the planets. The position of a planet with respect to the stars changes from night to night, sometimes exhibiting a “loop” motion, as in Fig. 2.1. The word “planet” is from the Greek “planetai”: wanderers.

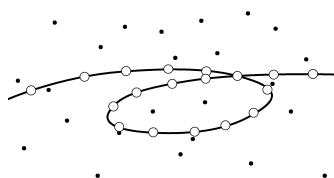


Fig. 2.1: The position of a planet ($\circ$) with respect to the stars changes nightly.

In the second century, Claudius Ptolemy devised a scheme to explain these motions. Ptolemy placed the Earth near the center of the universe with a planet moving on a small circle, called an *epicycle*, while the center of the epicycle moves (not at a uniform speed) on a larger circle, the *deferent*, around the Earth. See Fig. 2.2. By appropriately choosing the radii of the epicycle and deferent, as well as the speeds involved, Ptolemy was able to reproduce, with fair accuracy, the motions of the planets. This remarkable but cumbersome theory was accepted for over 1000 years.

In 1543 Nicholas Copernicus published a theory which was to revolutionize science and our perception of our place in the universe: he placed the Sun near the center of our planetary system, with the planets orbiting the Sun in circles at uniform speed.

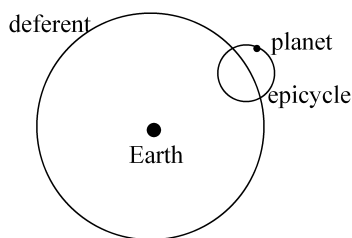


Fig. 2.2: Ptolemy's theory of planetary motion.

In 1609 Johannes Kepler published a theory which is more accurate than Copernicus': the path of a planet is an ellipse with the Sun at one focus. At about the same time Galileo Galilei was investigating the acceleration of objects near the Earth's surface. He found two things of interest for us: the acceleration is *constant* in time and *independent* of the mass and composition of the falling object.

In 1687 Isaac Newton published a theory of gravity which explained Kepler's astronomical and Galileo's terrestrial findings as manifestations of the same phenomenon: gravity. To understand how orbital motion is related to falling motion, refer to Fig. 2.3. The curves A, B, C are the paths of objects leaving the top of a tower with greater and greater horizontal velocities. They hit the ground farther and farther from the bottom of the tower until C when the object goes into orbit.

Mathematically, Newton's theory says that a planet in the Sun's gravity or an apple in the Earth's gravity is pulled instantaneously by the central body (do not ask how!), causing an acceleration

$$a = -\frac{\kappa M}{r^2}, \quad (2.1)$$

where κ is the *Newtonian gravitational constant*, M is the mass of the central body, and r is the distance to the center of the central body. Eq. (2.1) implies that the planets orbit the Sun in ellipses, in accord with Kepler's findings. See Appendix 7. By taking the distance r to the Earth's center to be sensibly constant near the Earth's surface, we see that Eq. (2.1) is also in accord with Galileo's findings: a is constant in time and is independent of the mass and composition of the falling object.

Newton's theory has enjoyed enormous success. A spectacular example occurred in 1846. Observations of the position of the planet Uranus disagreed with the predictions of Newton's theory of gravity, even after taking into account the gravitational effects of the other known planets. The discrepancy was about 4 arcminutes – $1/8^{\text{th}}$ of the angular diameter of the moon. U. Le Verrier, a French mathematician, calculated that a new planet, beyond Uranus, could account for the discrepancy. He wrote J. Galle, an astronomer at the Berlin observatory,

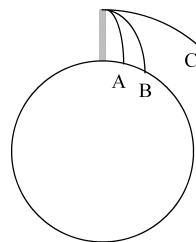


Fig. 2.3: Falling and orbital motion are the same.

telling him where the new planet should be – and Neptune was discovered! It was within 1 arcdegree of Le Verrier’s prediction.

Even today, calculations of spacecraft trajectories are made using Newton’s theory. The incredible accuracy of his theory will be examined further in Sec. [3.3](#).

Nevertheless, Einstein rejected Newton’s theory because it is based on pre-relativity ideas about time and space which, as we have seen, are not correct. For example, the acceleration in Eq. [\(2.1\)](#) is instantaneous with respect to a universal time.

2.2 The Key to General Relativity

A curved surface is different from a flat surface. However, a simple observation by the nineteenth century mathematician Karl Friedrich Gauss provides the key to the construction of the theory of surfaces: *a small region of a curved surface is very much like a small region of a flat surface*. This is familiar: a small region of a (perfectly) spherical Earth appears flat. This is why it took so long to discover that it is not flat. On an apple, a smaller region must be chosen before it appears flat.

In the next three sections we shall formalize Gauss' observation by taking the three postulates for flat surfaces from Chapter 1, restricting them to small regions, and then using them as postulates for curved surfaces.

We shall see that a curved spacetime is different from a flat spacetime. However, a simple observation of Einstein provides the key to the construction of general relativity: *a small region of a curved spacetime is very much like a small region of a flat spacetime*. To understand this, we must extend the concept of an inertial object to curved spacetimes.

Passengers in an airplane at rest on the ground or flying in a straight line at constant speed feel gravity, which feels very much like an inertial force. Accelerometers respond to the gravity. On the other hand, astronauts in orbit or falling radially toward the Earth feel no inertial forces, even though they are not moving in a straight line at constant speed with respect to the Earth. And an accelerometer carried by the astronauts will register zero. We shall include gravity as an inertial force and, as in special relativity, call an object *inertial* if it experiences no inertial forces. An inertial object in gravity is in free fall. Forces other than gravity do not act on it.

We now rephrase key Einstein's observation: *as viewed by inertial observers*, a small region of a curved spacetime is very much like a small region of a flat spacetime. We see this vividly in motion pictures of astronauts in orbit. No gravity is apparent in their cabin; objects suspended at rest remain at rest. Inertial objects in the cabin move in a straight line at constant speed, just as in a flat spacetime. The Newtonian theory predicts this: according to Eq. (2.1) an inertial object and the cabin accelerate the same with respect to the Earth and so they do not accelerate with respect to each other.

In the next three sections we shall formalize Einstein's observation by taking our three postulates for flat spacetimes, restricting them to small spacetime regions, and then using them as our first three (of four) postulates for curved spacetimes. The *local inertial frame postulate* asserts the existence of small inertial cubical lattices with synchronized clocks to serve as coordinate systems in small regions of a curved spacetime. The *metric postulate* asserts a universal light speed and a slowing of moving clocks in local inertial frames. The *geodesic postulate* asserts that inertial particles and light move in a straight line at constant speed in local inertial frames. We first discuss experimental evidence for the three postulates.

Experiments of R. Dicke and of V. B. Braginsky, performed in the 1960's, verify to extraordinary accuracy Galileo's finding incorporated into Newton's theory: the acceleration of a free falling object in gravity is independent of its mass and composition. (See Sec. 2.1.) We may reformulate this in the language of spacetimes: the worldline of an inertial object in a curved spacetime is independent of its mass and composition. The geodesic postulate will incorporate this by not referring to the mass or composition of the inertial objects whose worldlines it describes.

Dicke and Braginsky used the Sun's gravity. We can understand the principle of their experiments from the simplified diagram in Fig. 2.4. The weights A and B, supported by a quartz fiber, are, with the Earth, in free fall around the Sun. Dicke and Braginsky used various substances with various properties for the weights. Any difference in their acceleration toward the Sun would cause a twisting of the fiber. Due to the Earth's rotation, the twisting would be in the opposite direction twelve hours later. The apparatus had a resonant period of oscillation of 24 hours so that oscillations could build up. In Braginsky's experiment the difference in the acceleration of the weights toward the Sun was no more than one part in 10^{12} of their mutual acceleration toward the Sun. A planned satellite test (STEP) will test the equality of accelerations to one part in 10^{18} .

A related experiment shows that the Earth and the Moon, despite the huge difference in their masses, accelerate the same in the Sun's gravity. If this were not so, then there would be unexpected changes in the Earth-Moon distance. Changes in this distance can be measured within 2 cm (!) by timing the return of a laser pulse sent from Earth to mirrors on the Moon left by astronauts. This is part of the *lunar laser experiment*. The measurements show that the relative acceleration of the Earth and Moon is no more than a part in 10^{13} of their mutual acceleration toward the Sun. The experiment also shows that the Newtonian gravitational constant κ in Eq. (2.1) does not change by more than 1 part in 10^{12} per year. The constant also appears in Einstein's field equation, Eq. (2.27).

There is another difference between the Earth and the Moon which might cause a difference in their acceleration toward the Sun. Imagine disassembling to Earth into small pieces and separating the pieces far apart. The separation requires energy input to counteract the gravitational attraction of the pieces. This energy is called the Earth's *gravitational binding energy*. By Einstein's principle of the equivalence of mass and energy ($E = mc^2$), this energy is equivalent to mass. Thus the separated pieces have more total mass than the Earth. The difference is small – 4.6 parts in 10^{10} . But it is 23 times smaller for the Moon. One can wonder whether this difference between the Earth and

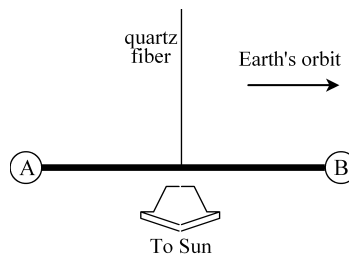


Fig. 2.4: Masses A and B accelerate the same toward the Sun.

Moon causes a difference in their acceleration toward the Sun. The lunar laser experiment shows that this does not happen. This is something that the Dicke and Braginsky experiments cannot test.

The last experiment we shall consider as evidence for the three postulates is the *terrestrial redshift experiment*. It was first performed by R. V. Pound and G. A. Rebka in 1960 and then more accurately by Pound and J. L. Snider in 1964. The experimenters put a source of gamma radiation at the bottom of a tower. Radiation received at the top of the tower was redshifted: $z = 2.5 \times 10^{-15}$, within an experimental error of about 1%. This is a *gravitational redshift*.

According to the discussion following Eq. (1.6), an observer at the top of the tower would see a clock at the bottom run slowly. Clocks at rest at different heights in the Earth's gravity run at different rates! Part of the result of the Hafele-Keating experiment is due to this. See Exercise 2.1.

We showed in Sec. 1.3 that the assumption Eq. (1.4), necessary for synchronizing clocks at rest in the coordinate lattice of an inertial frame, is equivalent to a zero redshift between the clocks. This assumption fails for clocks at the top and bottom of the tower. Thus clocks at rest in a small coordinate lattice on the ground cannot be (exactly) synchronized.

We now show that the experiment provides evidence that clocks at rest in a small inertial lattice *can* be synchronized. In the experiment, the tower has (upward) acceleration g , the acceleration of Earth's gravity, in a small inertial lattice falling radially toward Earth. We will show shortly that the same redshift would be observed with a tower having acceleration g in an inertial frame in a flat spacetime. This is another example of small regions of flat and curved spacetimes being alike. Thus it is reasonable to assume that there would be no redshift with a tower at rest in a small inertial lattice in gravity, just as with a tower at rest in an inertial frame. (It is desirable to test this directly by performing the experiment in orbit.) In this way, the experiment provides evidence that the condition Eq. (1.4), necessary for clock synchronization, is valid for clocks at rest in a small inertial lattice. Loosely speaking, we may say that since light behaves “properly” in a small inertial lattice, light accelerates the same as matter in gravity.

We now calculate the Doppler redshift for a tower with acceleration g in an inertial frame. Suppose the tower is momentarily at rest when gamma radiation is emitted. The radiation travels a distance h , the height of the tower, in the inertial frame. (We ignore the small distance the tower moves during the flight of the radiation. We shall also ignore the time dilation of clocks in the moving tower and the length contraction – see Appendix 5 – of the tower. These effects are far too small to be detected by the experiment.) Thus the radiation takes time $t = h/c$ to reach the top of the tower. (For clarity we do not take $c = 1$.) In this time the tower acquires a speed $v = gt = gh/c$ in the inertial frame. From Exercise 1.6, this speed causes a Doppler redshift

$$z = \frac{v}{c} = \frac{gh}{c^2}. \quad (2.2)$$

In the experiment, $h = 2250$ cm. Substituting numerical values in Eq.

(2.2) gives the value of z measured in the terrestrial redshift experiment; the gravitational redshift for towers accelerating in inertial frames is the same as the Doppler redshift for towers accelerating in small inertial lattices near Earth. Exercise 3.4 shows that a rigorous calculation in general relativity also gives Eq. (2.2).

Exercise 2.1. Let h be the height at which the airplane flies in the simplified Hafele-Keating experiment of Exercise 1.10. Show that the difference between the clocks due to the gravitational redshift is

$$\Delta s_a - \Delta s_r = gh\Delta t.$$

Suppose that $h = 10$ km. Substitute values to obtain $\Delta s_a - \Delta s_r = 1.6 \times 10^{-7}$ sec.

Adding this to the time dilation difference of Exercise 1.10 gives a difference of 3.0×10^{-7} sec. Exercise 3.3 shows that a rigorous calculation in general relativity gives the same result.

The Global Positioning System (GPS) of satellites must adjust its clocks for time dilation and gravitational redshifts to function properly. In fact, the effects are 10,000 times too large to be ignored.

2.3 The Local Inertial Frame Postulate

Suppose that curved surface dwellers attempt to construct a square coordinate grid using rigid rods constrained (of course) to their surface. If the rods are short enough, then at first they will fit together well. But, owing to the curvature of the surface, as the grid gets larger the rods must be forced a bit to connect them. This will cause stresses in the lattice and it will not be quite square. Surface dwellers call a small (nearly) square coordinate grid a *local planar frame* at P , where P is the point at the origin of the grid. In smaller regions around P , the grid must become more square. See Fig. 2.5.

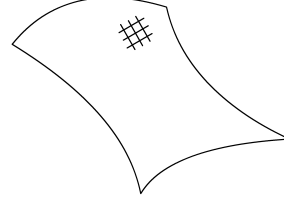


Fig. 2.5: A local planar frame.

The Local Planar Frame Postulate for a Curved Surface

A local planar frame can be constructed at any point P of a curved surface with any orientation.

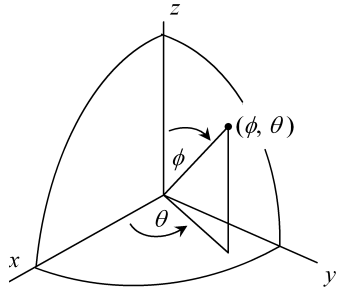


Fig. 2.6: Spherical coordinates (ϕ, θ) on a sphere.

We shall see that local planar frames at P provide surface dwellers with an intuitive description of properties of a curved surface at P . However, in order to study the surface as a whole, they need global coordinates, defined over the entire surface. There are, in general, no natural global coordinate systems to single out in a curved surface as planar frames are singled out in a flat surface. Thus they attach global coordinates (y^1, y^2) in an *arbitrary* manner. The only restrictions are that different points must have different coordinates and nearby

points must receive nearby coordinates. In general, the coordinates will not have a geometric meaning; they merely serve to label the points of the surface.

One common way for us (but not surface dwellers) to attach global coordinates to a curved surface is to parameterize it in three dimensional space:

$$x = x(y^1, y^2), \quad y = y(y^1, y^2), \quad z = z(y^1, y^2). \quad (2.3)$$

As (y^1, y^2) varies, (x, y, z) varies on the surface. Assign coordinates (y^1, y^2) to the point (x, y, z) on the surface given by Eq. (2.3). For example, Fig. 2.6 shows spherical coordinates $(y^1, y^2) = (\phi, \theta)$ on a sphere of radius R . The coordinates are assigned by the usual parameterization

$$x = R \sin \phi \cos \theta, \quad y = R \sin \phi \sin \theta, \quad z = R \cos \phi. \quad (2.4)$$

The Global Coordinate Postulate for a Curved Surface

The points of a curved surface can be labeled with coordinates (y^1, y^2) .

(Technically, the postulate should state that a curved surface is a two dimensional *manifold*. The statement given will suffice for us.)

In the last section we saw that inertial objects in an astronaut's cabin behave as if no gravity were present. Actually, they will not behave *exactly* as if no gravity were present. To see this, assume for simplicity that their cabin is falling radially toward Earth. Inertial objects in the cabin do not accelerate exactly the same with respect to the Earth because they are at slightly different distances and directions from the Earth's center. See Fig. 2.7. Thus, an object initially at rest near the top of the cabin will slowly separate from one initially at rest near the bottom. In addition, two objects initially at rest at the same height will slowly move toward each other as they both fall toward the center of the Earth. These changes in velocity are called *tidal accelerations*. (Why?) They are caused by small differences in the Earth's gravity at different places in the cabin. They become smaller in smaller regions of space and time, i.e., in smaller regions of spacetime.

Suppose that astronauts in a curved spacetime attempt to construct an inertial cubical lattice using rigid rods. If the rods are short enough, then at first they will fit together well. But as the grid gets larger, the lattice will have to resist tidal accelerations, and the rods cannot all be inertial. This will cause stresses in the lattice and it will not be quite cubical.

In the last section we saw that the terrestrial redshift experiment provides evidence that clocks in a small inertial lattice can be synchronized. Actually, due to small differences in the gravity at different places in the lattice, an attempt to synchronize the clocks with the one at the origin with the procedure of Sec. 1.3 will not quite work.

However, we can hope that the procedure will work with as small an error as desired by restricting the lattice to a small enough region of a spacetime.

A small (nearly) cubical lattice with (nearly) synchronized clocks is called a *local inertial frame* at E , where E is the event at the origin of the lattice when the clock there reads zero. In smaller regions around E , the lattice is more cubical and the clocks are more nearly synchronized. Local inertial frames are in free fall.

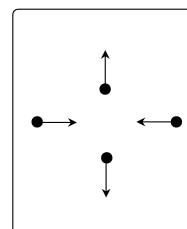


Fig. 2.7: Tidal accelerations in a radially free falling cabin.

The Local Inertial Frame Postulate for a Curved Spacetime

A local inertial frame can be constructed at any event E of a curved spacetime, with any orientation, and with any inertial object at E instantaneously at rest in it.

We shall find that local inertial frames at E provide an intuitive description of properties of a curved spacetime at E . However, in order to study a curved spacetime as a *whole*, we need global coordinates, defined over the entire spacetime. There are, in general, no natural global coordinates to single out in a curved spacetime, as inertial frames were singled out in a flat spacetime. Thus we attach global coordinates in an arbitrary manner. The only restrictions are that different events must receive different coordinates and nearby events must receive nearby coordinates. In general, the coordinates will not have a physical meaning; they merely serve to label the events of the spacetime.

Often one of the coordinates is a “time” coordinate and the other three are “space” coordinates, but this is not necessary. For example, in a flat spacetime it is sometimes useful to replace the coordinates (t, x) with $(t + x, t - x)$.

The Global Coordinate Postulate for a Curved Spacetime

The events of a curved spacetime can be labeled with coordinates (y^0, y^1, y^2, y^3) .

In the next two sections we give the metric and geodesic postulates of general relativity. We first express the postulates in local inertial frames. This *local form* of the postulates gives them the same physical meaning as in special relativity. We then translate the postulates to global coordinates. This *global form* of the postulates is unintuitive and complicated but is necessary to carry out calculations in the theory.

We can use arbitrary global coordinates in flat as well as curved spacetimes. We can then put the metric and geodesic postulates of special relativity in the same global form that we shall obtain for these postulates for curved spacetimes. We do not usually use arbitrary coordinates in flat spacetimes because inertial frames are so much easier to use. We do not have this luxury in curved spacetimes.

It is remarkable that we shall be able to describe curved spacetimes *intrinsically*, i.e., without describing them as curved in a higher dimensional flat space. Gauss created the mathematics necessary to describe curved surfaces intrinsically in 1827. G. B. Riemann generalized Gauss’ mathematics to curved spaces of higher dimension in 1854. His work was extended by several mathematicians. Thus the mathematics necessary to describe curved spacetimes intrinsically was waiting for Einstein when he needed it.

2.4 The Metric Postulate

Fig. 2.5 shows that local planar frames provide curved surface dwellers with a convenient way to express infinitesimal distances on a curved surface.

The Metric Postulate for a Curved Surface, Local Form

Let point Q have coordinates (dx^1, dx^2) in a local planar frame at P . Let ds be the distance between the points. Then

$$ds^2 = (dx^1)^2 + (dx^2)^2. \quad (2.5)$$

Even though the local planar frame extends a finite distance from P , Eq. (2.5) holds only for infinitesimal distances from P .

We now express Eq. (2.5) in terms of global coordinates. Set the matrix

$$\mathbf{f}^\circ = (f_{mn}^\circ) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then Eq. (2.5) can be written

$$ds^2 = \sum_{m,n=1}^2 f_{mn}^\circ dx^m dx^n. \quad (2.6)$$

Henceforth we use the *Einstein summation convention* by which an index which appears twice in a term is summed without using a Σ . Thus Eq. (2.6) becomes

$$ds^2 = f_{mn}^\circ dx^m dx^n. \quad (2.7)$$

As another example of the summation convention, consider a function $f(y^1, y^2)$. Then we may write the differential $df = \sum_{i=1}^2 (\partial f / \partial y^i) dy^i = (\partial f / \partial y^i) dy^i$.

Let P and Q be neighboring points on a curved surface with coordinates (y^1, y^2) and $(y^1 + dy^1, y^2 + dy^2)$ in a global coordinate system. Let Q have coordinates (dx^1, dx^2) in a local planar frame at P . We may think of the (x^i) coordinates as functions of the (y^j) coordinates, just as cartesian coordinates in the plane are functions of polar coordinates: $x = r \cos \theta$, $y = r \sin \theta$. This gives meaning to the partial derivatives $\partial x^i / \partial y^j$. From Eq. (2.7), the distance from P to Q is

$$\begin{aligned} ds^2 &= f_{mn}^\circ dx^m dx^n \\ &= f_{mn}^\circ \left(\frac{\partial x^m}{\partial y^j} dy^j \right) \left(\frac{\partial x^n}{\partial y^k} dy^k \right) \text{ (sum on } m, n, j, k) \\ &= \left(f_{mn}^\circ \frac{\partial x^m}{\partial y^j} \frac{\partial x^n}{\partial y^k} \right) dy^j dy^k \\ &= g_{jk}(y) dy^j dy^k, \end{aligned} \quad (2.8)$$

where we have set

$$g_{jk}(y) = f_{mn}^{\circ} \frac{\partial x^m}{\partial y^j} \frac{\partial x^n}{\partial y^k}. \quad (2.9)$$

Since (f_{mn}°) is symmetric, so is (g_{jk}) . Use a local planar frame at each point of the surface in this manner to translate the local form of the metric postulate, Eq. (2.5), to global coordinates:

The Metric Postulate for a Curved Surface, Global Form

Let (y^i) be global coordinates on the surface. Let ds be the distance between neighboring points (y^i) and $(y^i + dy^i)$. Then there is a symmetric matrix $(g_{jk}(y^i))$ such that

$$ds^2 = g_{jk}(y^i) dy^j dy^k. \quad (2.10)$$

$(g_{jk}(y^i))$ is called *metric* of the surface with respect to the coordinates (y^i) .

Exercise 2.2. Show that the metric for the (ϕ, θ) coordinates on the sphere in Eq. (2.4) is $ds^2 = R^2 d\phi^2 + R^2 \sin^2 \phi d\theta^2$, i.e.,

$$g(\phi, \theta) = \begin{pmatrix} R^2 & 0 \\ 0 & R^2 \sin^2 \phi \end{pmatrix}. \quad (2.11)$$

Do this in two ways:

a. By converting from the metric of a local planar frame. Show that for a local planar frame whose x^1 -axis coincides with a circle of latitude, $dx^1 = R \sin \phi d\theta$ and $dx^2 = R d\phi$. See Fig. 2.10.

b. Use Eq. (2.4) to convert $ds^2 = dx^2 + dy^2 + dz^2$ to (ϕ, θ) coordinates.

Exercise 2.3. Consider the hemisphere $z = (R^2 - x^2 - y^2)^{\frac{1}{2}}$. Assign coordinates (x, y) to the point (x, y, z) on the hemisphere. Find the metric in this coordinate system. Express your answer as a matrix. Hint: Use $z^2 = R^2 - x^2 - y^2$ to compute dz^2 .

We should not think of a vector as its components (v_i) , but as a single object $\mathbf{v}$ which represents a magnitude and direction (an arrow). In a given coordinate system the vector acquires components. The components will be different in different coordinate systems.

Similarly, we should not think of the metric as its components (g_{jk}) , but as a single object $\mathbf{g}$ which represents infinitesimal distances. In a given coordinate system the metric acquires components. The components will be different in different coordinate systems.

Exercise 2.4. Let (y^i) and $(\bar{y}^i)$ be two coordinate systems on the same surface, with metrics $(g_{jk}(y^i))$ and $(\bar{g}_{pq}(\bar{y}^i))$. Show that

$$\bar{g}_{pq} = g_{jk} \frac{\partial y^j}{\partial \bar{y}^p} \frac{\partial y^k}{\partial \bar{y}^q}. \quad (2.12)$$

Hint: See Eq. (2.8).

The Metric Postulate for a Curved Spacetime, Local Form

Let event F have coordinates (dx^i) in a local inertial frame at E . If E and F are lightlike separated, let $ds = 0$. If the events are timelike separated, let ds be the time between them as measured by any (inertial or noninertial) clock. Then

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \quad (2.13)$$

The metric postulate asserts *a universal light speed and a formula for proper time in local inertial frames*. (See the discussion following Eq. (1.11).) This is another instance of the key to general relativity: a small region of a curved spacetime is very much like a small region of a flat spacetime.

We now translate the metric postulate to global coordinates. Eq. (2.13) can be written

$$ds^2 = f_{mn}^\circ dx^m dx^n, \quad (2.14)$$

where

$$\mathbf{f}^\circ = (f_{mn}^\circ) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Using Eq. (2.14), the calculation Eq. (2.8), which produced the global form of the metric postulate for curved surfaces, now produces the global form of the metric postulate for curved spacetimes.

The Metric Postulate for a Curved Spacetime, Global Form

Let (y^i) be global coordinates on the spacetime. If neighboring events (y^i) and $(y^i + dy^i)$ are lightlike separated, let $ds = 0$. If the events are timelike separated, let ds be the time between them as measured by a clock moving between them. Then there is a symmetric matrix $(g_{jk}(y^i))$ such that

$$ds^2 = g_{jk}(y^i) dy^j dy^k. \quad (2.15)$$

$(g_{jk}(y^i))$ is called the *metric* of the spacetime with respect to (y^i) .

2.5 The Geodesic Postulate

Curved surface dwellers find that some curves in their surface are straight in local planar frames. They call these curves *geodesics*. Fig. 2.8 shows that the equator is a geodesic but the other circles of latitude are not. To traverse a geodesic, a surface dweller need only always walk “straight ahead”. A geodesic is as straight as possible, given that it is constrained to the surface.

As with the geodesic postulate for flat surfaces Eq. (1.15), we have

The Geodesic Postulate for a Curved Surface, Local Form

Parameterize a geodesic $x^i(s)$, where s is arclength. Let point P be on the geodesic. Then in every local planar frame at P

$$\ddot{x}^i(P) = 0, \quad i = 1, 2. \quad (2.16)$$

We now translate Eq. (2.16) to global coordinates y to obtain the global form of the geodesic equations. We first need to know that the metric $\mathbf{g} = (g_{ij})$ has an inverse $\mathbf{g}^{-1} = (g^{jk})$.

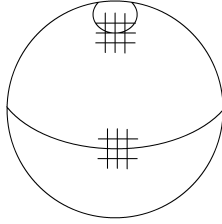


Fig. 2.8: The equator is the only circle of latitude which is a geodesic.

Exercise 2.5. a. Let the matrix $\mathbf{a} = (\partial x^n / \partial y^k)$. Show that the inverse matrix $\mathbf{a}^{-1} = (\partial y^k / \partial x^j)$.

b. Show that Eq. (2.9) can be written $\mathbf{g} = \mathbf{a}^t \mathbf{f} \mathbf{a}$, where t means “transpose”.

c. Prove that $\mathbf{g}^{-1} = \mathbf{a}^{-1} (\mathbf{f}^\circ)^{-1} (\mathbf{a}^{-1})^t$.

Introduce the notation $\partial_k g_{im} = \partial g_{im} / \partial y^k$. Define the *Christoffel symbols*:

$$\Gamma_{jk}^i = \frac{1}{2} g^{im} [\partial_k g_{jm} + \partial_j g_{mk} - \partial_m g_{jk}]. \quad (2.17)$$

Note that $\Gamma_{jk}^i = \Gamma_{kj}^i$. The Γ_{jk}^i , like the g_{jk} , are functions of the coordinates.

Exercise 2.6. Show that for the metric Eq.

(2.11)

$$\Gamma_{\theta\theta}^\phi = -\sin\phi \cos\phi \quad \text{and} \quad \Gamma_{\theta\phi}^\theta = \Gamma_{\phi\theta}^\theta = \cot\phi.$$

The remaining Christoffel symbols are zero.

You should not try to assign a geometric meaning to the Christoffel symbols; simply think of them as what appears when the geodesic equations are translated from their local form Eq. (2.16) (which have an evident geometric meaning) to their global form (which do not):

The Geodesic Postulate for a Curved Surface, Global Form

Parameterize a geodesic with arclength s . Then in every global coordinate system

$$\ddot{y}^i + \Gamma_{jk}^i \dot{y}^j \dot{y}^k = 0, \quad i = 1, 2. \quad (2.18)$$

Appendix 9 translates the local form of the postulate to the global form. The translation requires an assumption. A local planar frame at P extends to a finite region around P . Let $\mathbf{f} = (f_{mn}(x))$ represent the metric in this coordinate system. According to Eq. (2.7), $(f_{mn}(P)) = \mathbf{f}^\circ$. Appendix 8 shows that there are coordinates, called *geodesic coordinates*, satisfying this relationship and also

$$\partial_i f_{mn}(P) = 0 \quad (2.19)$$

for all m, n, i . A function with a zero derivative at a point is not changing much at the point. In this sense Eq. (2.19) states that $\mathbf{f}$ stays close to $\mathbf{f}^\circ$ near P . Since a local planar frame at P is constructed to approximate a planar frame as closely as possible near P , surface dwellers assume that the metric of a local planar frame at P satisfies Eq. (2.19).

Exercise 2.7. Show that the metric of Exercise 2.3 satisfies Eq. (2.19) at $(x, y) = (0, 0)$.

Exercise 2.8. Show that Eq. (2.18) reduces to Eq. (2.16) for local planar frames.

Exercise 2.9. Show that the equator is the only circle of latitude which is a geodesic. Of course, all great circles on a sphere are geodesics. Use the result of Exercise 2.6. Before using the geodesic equations you must parameterize the circles with s .

The Geodesic Postulate for a Curved Spacetime, Local Form

Worldlines of inertial particles and pulses of light can be parameterized so that if E is on the worldline, then in every local inertial frame at E

$$\ddot{x}^i(E) = 0, \quad i = 0, 1, 2, 3. \quad (2.20)$$

For inertial particles we may take the parameter to be s .

The worldlines are called *geodesics*.

The geodesic postulate asserts that *in a local inertial frame inertial particles and light move in a straight line at constant speed*. (See the remarks following Eq. (1.18).)

The worldline of an inertial particle or pulse of light in a curved spacetime is as straight as possible (in space *and* time – see the remarks following Eq. (1.17)), given that it is constrained to the spacetime. The geodesic is straight in local inertial frames, but looks curved when viewed in an “inappropriate” coordinate system. See Fig. 2.9.

Einstein’s “straightest worldline in a curved spacetime” description is very different from Newton’s “curved path in a flat space” description.

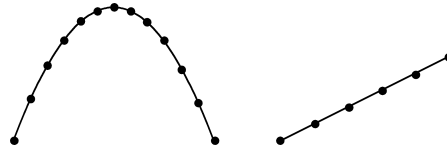


Fig. 2.9: The path of an inertial particle in a lattice stuck to the Earth and in an inertial lattice. The dots are at equal time intervals.

We now assume that the metric of a local inertial frame at E satisfies Eq. (2.19) for the same reasons as given above for local planar frames. Then Appendix 9 translates the local form of the geodesic postulate for curved spacetimes to the global form:

The Geodesic Postulate for a Curved Spacetime, Global Form

Worldlines of inertial particles and pulses of light can be parameterized so that in every global coordinate system

$$\boxed{\ddot{y}^i + \Gamma_{jk}^i \dot{y}^j \dot{y}^k = 0, \quad i = 0, 1, 2, 3.} \quad (2.21)$$

For inertial particles we may take the parameter to be s .

2.6 The Field Equation

Previous sections of this chapter explored similarities between small regions of flat and curved surfaces and between small regions of flat and curved spacetimes. This section explores differences.

The local forms of our curved surface postulates show that in many ways a small region of a curved surface is like a small region of a flat surface. Surface dwellers might suppose that all differences between the regions vanish as the regions become smaller. This is not so. To see this, pass geodesics through a point P in every direction.

Exercise 2.10. Show that there is a unique geodesic through every point of a curved surface in every direction. Hint: Use a basic theorem on the existence and uniqueness of solutions of systems of differential equations.

Connect all the points at distance r from P along the geodesics, forming a “circle” C of radius r . Let $C(r)$ be the circumference of the circle. Define the *curvature* K of the surface at P :

$$K = \frac{3}{\pi} \lim_{r \rightarrow 0} \frac{2\pi r - C(r)}{r^3}. \quad (2.22)$$

Clearly, $K = 0$ for a flat surface. From Fig. 2.10, $C(r) < 2\pi r$ for a sphere and so $K \geq 0$. From Fig. 2.10, we find

$$C(r) = 2\pi R \sin \phi = 2\pi R \sin(r/R) = 2\pi R [r/R - (r/R)^3/6 + \dots].$$

A quick calculation shows that $K = 1/R^2$. The curvature is a difference between regions of a sphere and a flat surface which does not vanish as the regions become smaller.

The surface of revolution of Fig. 2.11 is a *pseudo-sphere*. The horizontal “circles of latitude” are concave inward and the vertical “lines of longitude” are concave outward. Thus $C(r) > 2\pi r$ and so $K \leq 0$. Exercise 2.14 shows that $K = -1/R^2$, where R is a constant. Despite the examples, in general K varies from point to point in a curved surface.

Distances and geodesics are involved in the definition of K . But distances determine geodesics: distances determine the metric, which determines the Christoffel symbols Eq. (2.17), which determine the geodesics Eq. (2.18). Thus we learn an important fact: distances determine K . Thus K is measurable by surface dwellers.

Exercise 2.11. Show that a map of a region of the Earth must distort distances. Take the Earth to be perfectly spherical. Make no calculations.

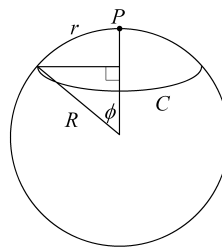


Fig. 2.10: $K = 1/R^2$ on a sphere.

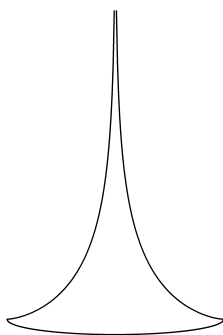


Fig. 2.11: $K = -1/R^2$ on a pseudo-sphere.

We can roll a flat piece of paper into a cylinder without distorting distances on the paper and thus without changing K . Thus $K = 0$ for the cylinder. Viewed from the outside, the cylinder is curved, and so $K = 0$ seems “wrong”. However, viewed from within (and remember, we are describing curved surfaces and spacetimes without reference to a higher dimensional space), the rolling does not distort distances in the paper. Thus surface dwellers could not detect the curvature seen from the outside. Thus K “should” be zero for the cylinder.

The formula expressing K in terms of distances was given by Gauss. If $g_{12} = 0$, then

$$K = -(g_{11}g_{22})^{-\frac{1}{2}} \left\{ \partial_1 \left(g_{11}^{-\frac{1}{2}} \partial_1 g_{22}^{\frac{1}{2}} \right) + \partial_2 \left(g_{22}^{-\frac{1}{2}} \partial_2 g_{11}^{\frac{1}{2}} \right) \right\} \quad (\partial_i \equiv \partial/\partial y^i). \quad (2.23)$$

Exercise 2.12. Show that Eq. (2.23) gives $K = 1/R^2$ for a sphere of radius R . Use Eq. (2.11).

Exercise 2.13. Generate a surface of revolution by rotating the parameterized curve $y = f(u), z = h(u)$ about the z -axis. Let (r, θ, z) be cylindrical coordinates and parameterize the surface with coordinates (u, θ) . Use $ds^2 = dr^2 + r^2 d\theta^2 + dz^2$ to show that the metric is

$$\begin{pmatrix} f'^2 + h'^2 & 0 \\ 0 & f^2 \end{pmatrix}.$$

Exercise 2.14. If $y = Re^{-u}$ and $z = R \int_0^u (1 - e^{-2t})^{\frac{1}{2}} dt$, then the surface of revolution in Exercise 2.13 is the pseudosphere of Fig. 2.11. Show that $K = -1/R^2$ for the pseudosphere.

Exercise 2.15. If $y = 1$ and $z = u$, then the surface of revolution in Exercise 2.13 is a cylinder. Show that $K = 0$ for a cylinder using Eq. (2.23).

The local forms of our curved spacetime postulates show that in many respects a small region of a curved spacetime is like a small region of a flat spacetime. We might suppose that all differences between the regions vanish as the regions become smaller. This is not so.

To see this, refer to Fig. 2.7. Let Δr be the small distance between objects at the top and bottom of the cabin and let Δa be the small tidal acceleration between them. In the curved spacetime of the cabin $\Delta a \neq 0$, which is different from the flat spacetime value $\Delta a = 0$. But in the cabin $\Delta a \rightarrow 0$ as $\Delta r \rightarrow 0$; this difference between a curved and flat spacetime *does* vanish as the regions become smaller. But also in the cabin $\Delta a/\Delta r \neq 0$, again different from the flat spacetime $\Delta a/\Delta r = 0$. This difference does *not* vanish as the regions become smaller: in the cabin, using Eq. (2.1), $\Delta a/\Delta r \rightarrow da/dr = 2\kappa M/r^3 \neq 0$.

The metric and geodesic postulates describe the behavior of clocks, light, and inertial particles in a curved (or flat) spacetime. But to apply these postulates, we must know the metric of the spacetime. Our final postulate for general relativity, the *field equation*, determines the metric. Loosely speaking, the equation determines the “shape” of a spacetime, how it is “curved”.

We constructed the metric in Sec. 2.4 using local inertial frames. There is obviously a relationship between the motion of local inertial frames and the distribution of mass in a curved spacetime. Thus, *there is a relationship between the metric of a spacetime and the distribution of mass in the spacetime*. The field equation gives this relationship. Schematically it reads

$$\boxed{\begin{bmatrix} \text{quantity determined} \\ \text{by metric} \end{bmatrix}} = \boxed{\begin{bmatrix} \text{quantity determined} \\ \text{by mass/energy} \end{bmatrix}}. \quad (2.24)$$

To specify the two sides of this equation, we need several definitions. Define the *Ricci tensor*

$$R_{jk} = \Gamma_{tk}^p \Gamma_{jp}^t - \Gamma_{tp}^p \Gamma_{jk}^t + \partial_k \Gamma_{jp}^p - \partial_p \Gamma_{jk}^p. \quad (2.25)$$

Don't panic over this convoluted definition: You need not have a physical understanding of the Ricci tensor. And while the R_{jk} are extremely tedious to calculate by hand, computers can readily calculate them for us.

The Ricci tensor is entirely determined by the metric. We shall see that it contains information about the curvature K of two dimensional surfaces in four dimensional spacetime. Like K , it involves second derivatives of the g_{jk} (because the Γ_{jk}^i involve the first derivatives).

As with the metric $\mathbf{g}$, we will use $\mathbf{R}$ to designate the Ricci tensor as a single object, existing independently of any coordinate system, but which in a given coordinate system acquires components R_{jk} .

Define the *curvature scalar* $R = g^{jk} R_{jk}$.

We can now specify the left side of the schematic field equation Eq. (2.24): the quantity determined by the metric is the *Einstein tensor*

$$\mathbf{G} = \mathbf{R} - \frac{1}{2} R \mathbf{g}.$$

The right side of the field equation is given by the *energy-momentum tensor* $\mathbf{T}$. It represents the source of the gravitational field in general relativity. All forms of matter and energy, including electromagnetic fields, and also pressures, stresses, and viscosity contribute to $\mathbf{T}$. But for our purposes we need to consider only matter of a special form, called *dust*. In dust, matter interacts only gravitationally; there are no pressures, stresses, or viscosity. Gas in interstellar space which is thin enough so that particle collisions are infrequent is dust.

We now define $\mathbf{T}$ for dust. Choose an event E with coordinates (y^i) . Let ρ be the density of the dust at E as measured by an observer moving with the dust. (Thus ρ is the same in all coordinate systems.) Let ds be the time measured by the observer between E and a neighboring event on the dust's worldline with coordinates $(y^i + dy^i)$. Define

$$\boxed{T^{jk} = \rho \frac{dy^j}{ds} \frac{dy^k}{ds}}. \quad (2.26)$$

We can now state our final postulate for general relativity:

The Field Equation

$$\boxed{\mathbf{G} = -8\pi\kappa\mathbf{T}.} \quad (2.27)$$

Here κ is the Newtonian gravitational constant of Eq. (2.1).

The field equation is the centerpiece of Einstein's theory. It relates, event by event, the curvature of a spacetime, represented by $\mathbf{G}$, to the density of matter and energy in the spacetime, represented by $\mathbf{T}$. In our applications, we will specify $\mathbf{T}$. The equation is then a system of second order *nonlinear* partial differential equations in the unknown $\mathbf{G}$.

The metric, the geodesic equations, and the field equation are valid in all coordinate systems. They are examples of our ability to describe curved spacetimes intrinsically, i.e., without reference to a higher dimensional flat space in which they are curved. You should be impressed with the power of this mathematics!

If $\rho = 0$ at some event, then the field equation is $R_{jk} - \frac{1}{2}Rg_{jk} = 0$. Multiply this by g^{jk} :

$$0 = g^{jk} (R_{jk} - \frac{1}{2}Rg_{jk}) = g^{jk}R_{jk} - \frac{1}{2}Rg_{jk}g^{jk} = R - \frac{1}{2}R4 = -R.$$

Substitute $R = 0$ into the field equation to obtain

The Vacuum Field Equation

$$\boxed{\mathbf{R} = \mathbf{0}.} \quad (2.28)$$

At events in a spacetime where there is no matter we may use this *vacuum field equation*.

Before we can use the field equation we must deal with a technical matter. The indices on the metric and the Ricci tensor are subscripts: g_{jk} and R_{jk} . The indices on the energy-momentum tensor are superscripts: T^{jk} . By convention, the placement of indices indicates how components transform under a change of coordinates.¹

Subscripted components a_{jk} transform *covariantly*:

$$\bar{a}_{pq} = a_{jk} \frac{\partial y^j}{\partial \bar{y}^p} \frac{\partial y^k}{\partial \bar{y}^q}. \quad (2.29)$$

According to Exercise 2.4, the g_{jk} transform covariantly. The R_{jk} also transform covariantly. (Do not attempt to verify this at home!) The Ricci scalar is the same in all coordinate systems. Thus the left side of the field equation, $G_{jk} = R_{jk} - \frac{1}{2}Rg_{jk}$, transforms covariantly.

¹This is a convention of *tensor algebra*. Tensor algebra and calculus are powerful tools for computations in general relativity. But we do not need them for a conceptual understanding of the theory. Appendix 10 is a short introduction to tensors.

Superscripted components a^{jk} transform *contravariantly*:

$$\bar{a}^{pq} = a^{jk} \frac{\partial \bar{y}^p}{\partial y^j} \frac{\partial \bar{y}^q}{\partial y^k}. \quad (2.30)$$

Exercise 2.16. Show that the T^{jk} transform contravariantly.

Since the G_{jk} and T^{jk} transform differently, we cannot take $G_{jk} = -8\pi\kappa T^{jk}$ as the components of the field equation: even if this equation were true in one coordinate system, it need not be in another. The solution is to *raise* the covariant indices to contravariant indices: $G^{mn} = g^{mj}g^{nk}G_{jk}$.

Exercise 2.17. Show that if the G_{jk} transform covariantly, then the G^{mn} transform contravariantly.

Then we can take the components of the field equation to be $G^{jk} = -8\pi\kappa T^{jk}$. If this equation is valid in any one coordinate system, then, since the two sides transform the same between coordinate systems, it is true in all. (Alternatively, we can *lower* the contravariant indices to covariant indices: $T_{mn} = g_{mj}g_{nk}T^{jk}$, and use the equivalent equation $G_{jk} = -8\pi\kappa T_{jk}$.)

Curvature. The field equation gives a simple and elegant relationship between the curvatures K of certain surfaces through an event and the density ρ at the event. To obtain it we must use Fermi normal coordinates, discussed in Appendix 11.

The metric $\mathbf{f}$ of a local inertial frame at E satisfies Eq. (2.14), $(f_{mn}(E)) = \mathbf{f}^\circ$. The metric of a geodesic coordinate system satisfies in addition Eq. (2.19), $\partial_i f_{mn}(E) = 0$. And in a spacetime, the metric of a *Fermi normal coordinate system* satisfies further $\partial_0 \partial_i f_{mn}(E) = 0$.

Consider the element of matter at an event E . According to the local inertial frame postulate, the element is instantaneously at rest in some local inertial frame at E , which we take to have Fermi normal coordinates. Let K_{12} be the curvature of the surface formed by holding the time coordinate x^0 and the spatial coordinate x^3 fixed, while varying the other two spatial coordinates, x^1 and x^2 . Appendix 11 shows that $G^{00} = -(K_{12} + K_{23} + K_{31})$. Thus from the field equation

$$K_{12} + K_{23} + K_{31} = 8\pi\kappa\rho.$$

In particular, if $\rho = 0$, then $K_{12} + K_{23} + K_{31} = 0$.

Why the Field Equation? We close this chapter with a plausibility argument, based on reasonable assumptions, which leads from the schematic field equation Eq. (2.24) to the field equation Eq. (2.27). This will by no means be a proof of the equation, but it should be convincing enough to make us anxious to confront the theory with experiment in the next chapter. The reader may turn to the next chapter without loss of continuity.

We take Eq. (2.24) to be a local equation, i.e., it will equate the two quantities event by event.

Consider first the right side of Eq. (2.24). It is reasonable to involve the density of matter. In special relativity there are two effects affecting the density

of moving matter. First, the mass of a body moving with speed v increases by a factor $(1 - v^2)^{-\frac{1}{2}}$. According to Eq. (1.12), this factor is dx^0/ds . Second, from Appendix 5, an inertial body contracts in its direction of motion by the same factor and does not contract in directions perpendicular to its direction of motion. Thus the density of moving matter is

$$T^{00} = \rho \frac{dx^0}{ds} \frac{dx^0}{ds}, \quad (2.31)$$

where ρ is the density measured by an observer moving with the matter. As emphasized in Chapter 1, the coordinate difference dx^0 has no physical significance in and of itself. Thus Eq. (2.31) is only one component of a whole:

$$T^{jk} = \rho \frac{dx^j}{ds} \frac{dx^k}{ds}. \quad (2.32)$$

This quantity represents matter in special relativity. A local inertial frame is in many respects like an inertial frame in special relativity. Thus at the origin of a local inertial frame we replace the right side of Eq. (2.24) with Eq. (2.32). Transforming to global coordinates, the right side of Eq. (2.24) is the energy-momentum tensor $\mathbf{T}$.

Since the T^{jk} transform contravariantly, the left side of Eq. (2.24) must transform contravariantly. We denote it G^{jk} in anticipation that it will turn out to be the Einstein tensor. Thus the field equation is of the form

$$G^{jk} = -8\pi\kappa T^{jk}. \quad (2.33)$$

(The factor $-8\pi\kappa$ was inserted for later convenience.)

We shall make four assumptions which will uniquely determine the G^{jk} .

(i) From Eq. (2.24), $\mathbf{G}$ depends on $\mathbf{g}$. Assume that the left side of Eq. (2.24), like the curvature K of a surface, depends only on the g_{jk} and their first and second derivatives.

(ii) In special relativity $\partial T^{jk}/\partial x^j = 0$. (For $k = 0$ this expresses conservation of mass and for $k = 1, 2, 3$ it expresses conservation of the k component of momentum.) Assume that $\partial T^{jk}/\partial x^j = 0$ at the origin of local inertial frames.

According to a mathematical theorem of Lovelock, our assumptions already imply that there is no loss of generality in taking the field equation to be of the form

$$A [\mathbf{R} - \frac{1}{2} \mathbf{g}R] + \Lambda \mathbf{g} = -8\pi\kappa \mathbf{T}, \quad (2.34)$$

where A and Λ are constants. See Appendix 12.

(iii) Assume that a spacetime without matter is flat. (We will revisit this assumption in Sec. 4.4.) In a flat spacetime the Christoffel symbols Eq. (2.17), the Ricci tensor Eq. (2.25), and the curvature scalar R all vanish. Thus Eq. (2.34) becomes $\Lambda \mathbf{g} = 0$. Thus $\Lambda = 0$.

(iv) Assume that Einstein's theory agrees with Newton's when Newton's is accurate, namely, for weak gravity with small velocities. Appendix 12 shows that this requires that $A = 1$. Thus Eq. (2.34) is precisely the field equation Eq. (2.27).

Chapter 3

Spherically Symmetric Spacetimes

3.1 Stellar Evolution

Several applications of general relativity described in this chapter involve observations of stars at various stages of their life. We thus begin with a brief survey of the relevant aspects of stellar evolution.

Clouds of interstellar gas and dust are a major component of our Milky Way galaxy. Suppose some perturbing force causes a cloud to begin to contract by self gravitation. If the mass of the cloud is small enough, then gas pressures and/or mechanical forces will stop the contraction and a planet sized object will result. For a larger cloud, these forces cannot stop the contraction. The cloud will continue to contract and become hotter until thermonuclear reactions begin. The heat from these reactions will increase the pressure in the cloud and stop the contraction. A star is born! Our own star, the Sun, formed in this way 4.6 billion years ago. A star will shine for millions or billions of years until its nuclear fuel runs out and it begins to cool. Then the contraction will begin again. The subsequent evolution of the star depends on its mass.

For a star of ~ 1.4 solar masses or less, the contraction will be stopped by a phenomenon known as *degenerate electron pressure* – but not until enormous densities are reached. For example, the radius of the Sun will decrease by a factor of 10^2 – to about 1000 km – and its density will thus increase by a factor of 10^6 , to about 10^6 g/cm^3 ! Such a star is a *white dwarf*. They are common. For example, Sirius, the brightest star in the sky, is part of a double star system. Its dim companion is a white dwarf. The Sun will become a white dwarf in about 5 billion years.

A white dwarf member of a binary star system often accretes matter from its companion. If it accretes sufficient matter, it will erupt in a thermonuclear explosion, called a Type Ia *supernova*, destroying the star. At its brightest, a Type Ia supernova has a luminosity over a billion times that of the Sun.

For a star over 1.4 solar masses, degenerate electron pressure cannot stop the contraction. Eventually a supernova of another kind, called Type II, occurs. The outer portions of the star blow into interstellar space. A supernova was recorded in China in 1054. It was visible during the day for 23 days and outshone all other stars in the sky for several weeks. Today we see the material blown from the star as the *Crab Nebula*.

One possible result of a Type II supernova is a *neutron star*. In a neutron star, fantastic pressures force most of the electrons to combine with protons to form neutrons. *Degenerate neutron pressure* prevents the star from collapsing further. A typical neutron star has a radius of 10 km and a density of 10^{14}g/cm^3 !

Neutron stars manifest themselves in two ways. They often spin rapidly – up to nearly 1000 times a second. For poorly understood reasons, there can be a small region of the star that emits radio and optical frequency radiation in a narrow cone. If the Earth should happen to be in the direction of the cone periodically as the star rotates, then the star will appear to pulse at the frequency of rotation of the star, rather like a lighthouse. The neutron star is a *pulsar*. The first pulsar was discovered in 1968; many are now known. There is a pulsar at the center of the crab nebula.

There are also neutron stars which emit X-rays. They are always members of a binary star system. We briefly discuss these systems in Sec. 3.6.

Even degenerate neutron pressure cannot stop the contraction if the star is too massive. A massive star can collapse to a *black hole*. Sec. 3.6 is devoted to black holes.

We close our catalog of remarkable astronomical objects with *quasars*, discovered in 1963. Quasars sit at the center of some galaxies, mostly distant. A typical quasar emits 100 times the energy of our entire Milky Way galaxy from a region 10^{17} times smaller!

3.2 The Schwarzschild Metric

In this section we solve the field equation for the metric of a spacetime around a spherically symmetric object such as the Sun. This will enable us in the next section to compare the predictions of general relativity with observations of the motion of planets and light in our solar system.

Exercise 3.1. Show that in spherical coordinates the flat spacetime metric Eq. (1.14) becomes

$$ds^2 = dt^2 - dr^2 - r^2 d\Omega^2, \quad (3.1)$$

where from Eq. (2.11), $d\Omega^2 = d\phi^2 + \sin^2\phi d\theta^2$ is the metric of the unit sphere. Hint: Don't calculate; think geometrically.

How can Eq. (3.1) change in a curved spherically symmetric spacetime? In such a spacetime:

- The θ and ϕ coordinates can have their usual meaning. (As in a curved surface, angles can be measured in the usual way in a curved spacetime.)
- $d\theta$ and $-d\theta$ produce the same ds , as do $d\phi$ and $-d\phi$. Thus none of the terms $drd\theta$, $drd\phi$, $d\theta d\phi$, $dt d\theta$, or $dt d\phi$ can appear in the metric.
- The surface $t = t_o$, $r = r_o$ has the metric of a sphere, although not necessarily of radius r_o .

Thus the metric is of the form

$$ds^2 = e^{2\mu} dt^2 - 2u dt dr - e^{2\nu} dr^2 - r^2 e^{2\lambda} d\Omega^2, \quad (3.2)$$

where μ, ν, u , and λ are unknown functions of t and r , but not, by symmetry, of θ or ϕ . We write some of the coefficients as exponentials for convenience.

The coordinate change $\bar{r} = r e^\lambda$ eliminates the $e^{2\lambda}$ factor. (Then rename the radial coordinate r .) A change in the t coordinate eliminates the $dt dr$ term. See Appendix 13. These coordinate changes put the metric in a simpler form:

$$ds^2 = e^{2\mu} dt^2 - e^{2\nu} dr^2 - r^2 d\Omega^2. \quad (3.3)$$

This is as far as we can go with symmetry and coordinate changes. We must use the field equation to determine μ and ν . Since we are interested only in the spacetime outside the central object, we use the vacuum field equation Eq. (2.28): $\mathbf{R} = \mathbf{0}$. The components of $\mathbf{R}$ are given by Eq. (2.25). They are best calculated with a computer.

From $R_{tr} = -2(\partial\nu/\partial t)/r = 0$, $\partial\nu/\partial t = 0$, i.e., ν depends only on r . Now

$$\begin{aligned} R_{tt} &= (-\mu'' + \mu'\nu' - \mu'^2 - 2\mu'r^{-1}) e^{2(\mu-\nu)} \\ R_{rr} &= (-\mu'' + \mu'\nu' - \mu'^2 + 2\nu'r^{-1}) \\ R_{\phi\phi} &= (-r\mu' + r\nu' + e^{2\nu} - 1) e^{-2\nu}, \end{aligned} \quad (3.4)$$

where primes indicate differentiation with respect to r .

Set $R_{tt} = 0$ and $R_{rr} = 0$, cancel the exponential factor, and subtract: $\mu' + \nu' = 0$. Now $R_{\phi\phi} = 0$ implies $2r\nu' + e^{2\nu} - 1 = 0$, or $(re^{-2\nu})' = 1$. Integrate: $e^{-2\nu} = 1 - 2m/r$, where $2m$ is a constant of integration. Since ν does not depend on t , neither does m . The metric Eq. (3.3) is now of the form

$$ds^2 = \left(1 - \frac{2m}{r}\right) e^{2(\mu+\nu)} dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2 d\Omega^2.$$

Since $\mu' + \nu' = 0$, $\mu + \nu$ is a function only of t . Thus the substitution $d\bar{t} = e^{\mu+\nu} dt$ eliminates the $e^{2(\mu+\nu)}$ factor. We arrive at our solution, the *Schwartzschild metric*, obtained by Karl Schwartzschild in 1916:

Schwartzschild Metric

$$ds^2 = \left(1 - \frac{2m}{r}\right) dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2 d\Omega^2. \quad (3.5)$$

We shall see later that

$$m = \kappa M, \quad (3.6)$$

where κ is the Newtonian gravitational constant of Eq. (2.1) and M is the mass of the central object. For the Sun,

$$m = 4.92 \times 10^{-6} \text{ sec} = 1.47 \text{ km}.$$

If $r \gg 2m$, then the Schwartzschild metric Eq. (3.5) and the flat spacetime metric Eq. (3.1) are nearly identical. Thus if $r \gg 2m$, then for most purposes r may be considered radial distance and t time measured by slowly moving clocks.

Exercise 3.2. Show that the time ds measured by a clock at rest at r is related to the coordinate time dt by

$$ds = \left(1 - \frac{2m}{r}\right)^{\frac{1}{2}} dt. \quad (3.7)$$

Exercise 3.3. Show that the Schwartzschild metric predicts the sum of the results of Exercises 1.10 (time dilation) and 2.1 (gravitational redshift) for the clock difference in Hafele-Keating experiment: $\Delta s_a - \Delta s_r = \left(\frac{1}{2}(v_a^2 - v_r^2) + gh\right) \Delta t$.

Gravitational Redshift. We investigate the gravitational redshift in a Schwartzschild spacetime. Emit pulses of light radially outward from (t_e, r_e) and $(t_e + \Delta t_e, r_e)$. Observe them at (t_o, r_o) and $(t_o + \Delta t_o, r_o)$. Since the Schwartzschild metric is time independent, the worldline of the second pulse is simply a translation of the first by Δt_e . Thus $\Delta t_o = \Delta t_e$. Let Δs_e be the time between the emission events as measured by a clock at r_e , with Δs_o defined similarly. By Eqs. (1.6) and (3.7) an observer at r_o finds a redshift

$$z = \frac{\Delta s_o}{\Delta s_e} - 1 = \frac{(1 - 2m/r_o)^{\frac{1}{2}} \Delta t_o}{(1 - 2m/r_e)^{\frac{1}{2}} \Delta t_e} - 1 = \left(\frac{1 - 2m/r_o}{1 - 2m/r_e}\right)^{\frac{1}{2}} - 1. \quad (3.8)$$

Exercise 3.4. Show that if $r_e \gg 2m$ and $r_o - r_e$ is small, then Eq. (3.8) reduces to Eq. (2.2), the gravitational redshift formula derived in connection with the terrestrial redshift experiment. Note that $c = 1$ in Eq. (3.8).

For light emitted at the surface of the Sun and received at Earth, Eq. (3.8) gives $z = 2 \times 10^{-6}$. This is difficult to measure but it has been verified within 7%. Light from a star with the mass of the Sun but a smaller radius will, by Eq. (3.8), have a larger redshift. For example, light from the white dwarf companion to Sirius has a gravitational redshift $z = 3 \times 10^{-4}$. And a gravitational redshift $z = .35$ has been measured in X-rays emitted from the surface of a neutron star.

Exercise 3.5. Show that a clock on the Sun will lose 63 sec/year compared to a clock far from the Sun. (The corresponding losses for the white dwarf and neutron star just mentioned are 2.6 hours and 95 days, respectively.)

An accurate measurement of the gravitational redshift was made in 1976 by an atomic clock in a rocket. The reading of the clock was compared, via radio, with one on the ground during the two hour flight of the rocket. Of course the gravitational redshift changed with the changing height of the rocket. After taking into account the redshift due to the motion of the rocket, the gravitational redshift predicted by general relativity was confirmed within 7 parts in 10^5 .

Geodesics. To determine the motion of inertial particles and light in a Schwarzschild spacetime, we must solve the geodesic equations Eq. (2.21). They are:

$$\ddot{t} + \frac{2m/r^2}{1 - 2m/r} \dot{r} \dot{t} = 0 \quad (3.9)$$

$$\ddot{r} + \frac{m(1 - 2m/r)}{r^2} \dot{t}^2 - \frac{m/r^2}{1 - 2m/r} \dot{r}^2 - r(1 - 2m/r) (\dot{\phi}^2 + \sin^2 \phi \dot{\theta}^2) = 0 \quad (3.10)$$

$$\ddot{\theta} + \frac{2}{r} \dot{r} \dot{\theta} + 2 \cot \phi \dot{\phi} \dot{\theta} = 0 \quad (3.11)$$

$$\ddot{\phi} + \frac{2}{r} \dot{r} \dot{\phi} - \sin \phi \cos \phi \dot{\theta}^2 = 0. \quad (3.12)$$

From spherical symmetry a geodesic lies in a plane. Let it be the plane

$$\phi = \pi/2. \quad (3.13)$$

This is a solution to Eq. (3.12).

Exercise 3.6. Use Eqs. (3.10) and (3.5) to show that $ds = (1 - 3m/r)^{\frac{1}{2}} dt$ for a circular orbit at $\phi = \pi/2$. (Cf. Exercise 3.2.) This shows that a clock (with $ds > 0$) can have a circular orbit only for $r > 3m$. It also shows that light (with $ds = 0$) can orbit at $r = 3m$. (Of course the central object must be inside the r of the orbit.)

Integrate Eqs. (3.9) and (3.11):

$$\dot{t} (1 - 2m/r) = B \quad (3.14)$$

$$\dot{\theta} r^2 = A, \quad (3.15)$$

where A and B are constants of integration. Physically, A and B are conserved quantities: A is the angular momentum per unit mass along the geodesic and B is the energy per unit mass.

Exercise 3.7. Differentiate Eqs. (3.14) and (3.15) to obtain Eqs. (3.9) and (3.11). (Remember that $\phi = \pi/2$.)

Substitute Eqs. (3.13)-(3.15) into Eq. (3.10) and integrate:

$$\frac{r^2}{1 - 2m/r} + \frac{A^2}{r^2} - \frac{B^2}{1 - 2m/r} = -E = \begin{cases} 0 & \text{for light} \\ -1 & \text{for inertial particles,} \end{cases} \quad (3.16)$$

where E is a constant of integration. The values given for E are verified by substituting Eqs. (3.13)-(3.16) into the Schwarzschild metric Eq. (3.5), giving $(ds/dp)^2 = E$. For light $ds = 0$ (see Eq. (2.16)) and so $E = 0$. For an inertial particle we may take $p = s$ (see Eq. (2.21)) and so $E = 1$.

Summary. The partially integrated geodesic equations Eqs. (3.13)-(3.16) will be the starting point for the study of the motion of planets and light in the solar system in the next section. It was a long journey to obtain the equations, so it might be useful to outline the steps used to obtain them:

- We obtained a general form for the metric of a spherically symmetric spacetime, Eq. (3.2).
- The Ricci tensor involves second derivatives of a metric. In Eqs. (3.4) we wrote some of the components of this tensor for the spherically symmetric metric.
- We wrote the vacuum field equations for the metric by setting the components of the Ricci tensor to zero.
- We solved the vacuum field equations to obtain the Schwarzschild metric, Eq. (3.5).
- The geodesic equations involve first derivatives of a metric. In Eqs. (3.9)-(3.12) we wrote the geodesic equations for the Schwarzschild metric. They are of second order.
- We integrated the second order geodesic equations to obtain first order Eqs. (3.13)-(3.16).

The Constant m . We close this section by evaluating the constant m in the Schwarzschild metric Eq. (3.5). Consider radial motion of an inertial particle. By Eq. (3.15), $A = 0$. Thus Eq. (3.16) becomes

$$\frac{dr}{ds} = \pm \left(B^2 - 1 + \frac{2m}{r} \right)^{\frac{1}{2}}, \quad (3.17)$$

the sign chosen according as the motion is outward or inward. Differentiate Eq. (3.17) and substitute Eq. (3.17) into the result:

$$\frac{d^2r}{ds^2} = -\frac{m}{r^2}. \quad (3.18)$$

For a distant slowly moving particle, $ds \approx dt$. Since the Newtonian theory applies to this situation, Eqs. (2.1) and (3.18) must coincide. Thus $m = \kappa M$, in agreement with Eq. (3.6).

3.3 The Solar System Tests

This section compares the predictions of general relativity with observations of the motion of planets and light in our solar system. Three general relativistic effects have been measured: perihelion advance, light deflection, and light delay.

Perihelion Advance. We first solve the geodesic equations for the planets. Set $u = 1/r$. By Eq. (3.15),

$$\dot{r} = \frac{dr}{du} \frac{du}{d\theta} \dot{\theta} = -u^{-2} \frac{du}{d\theta} A r^{-2} = -A \frac{du}{d\theta}.$$

Substitute this into Eq. (3.16), multiply by $1 - 2mu$, differentiate with respect to θ , and divide by $2A^2 du/d\theta$:

$$\frac{d^2 u}{d\theta^2} + u = mA^{-2}E + 3mu^2. \quad (3.19)$$

According to Eq. (3.16), $E = 1$ for the planets.

Exercise 3.8. Among the planets, Mercury has the largest ratio $3mu^2/mA^{-2}$ of the terms on the right side of Eq. (3.19). Show that it is less than 10^{-7} .

As a first approximation, solve Eq. (3.19) without the small $3mu^2$ term:

$$u = mA^{-2} [1 + e \cos(\theta - \theta_p)]. \quad (3.20)$$

This is the polar equation of an ellipse with eccentricity e and perihelion (point of closest approach to the Sun) at $\theta = \theta_p$. The same equation was derived from Newton's theory in Appendix 7. Thus, although Einstein's theory is conceptually entirely different from Newton's, it gives nearly the same predictions for the planets. This is necessary for any theory of gravity, as Newton's theory is very accurate for the planets.

Set $\theta = \theta_p$ in Eq. (3.20):

$$mA^{-2} = r_p^{-1} (1 + e)^{-1} \quad (3.21)$$

To obtain a more accurate solution of Eq. (3.19), substitute Eqs. (3.20) and (3.21) into the right side of Eq. (3.19) and solve:

$$u = r_p^{-1} (1 + e)^{-1} \left\{ 1 + e \left[\cos(\theta - \theta_p) + 3mr_p^{-1} (1 + e)^{-1} \theta \sin(\theta - \theta_p) \right] \right\} + mr_p^{-2} (1 + e)^{-2} \left\{ 3 + \frac{1}{2}e^2 [3 - \cos(2(\theta - \theta_p))] \right\}. \quad (3.22)$$

Drop the last term, which stays small because of the factor mr_p^{-2} . For small α ,

$$\cos \beta + \alpha \sin \beta \approx \cos \beta \cos \alpha + \sin \beta \sin \alpha = \cos(\beta - \alpha).$$

Use this approximation in Eq. (3.22):

$$u = r_p^{-1} (1 + e)^{-1} \left\{ 1 + e \cos \left[\theta - \left(\theta_p + \frac{3m\theta}{r_p(1+e)} \right) \right] \right\}. \quad (3.23)$$

Since $3m/r_p(1+e)$ is small, this may be considered the equation of an ellipse (c.f. Eq. (3.20)) with perihelion advanced by $6m\pi/r_p(1+e)$ per revolution.

Exercise 3.9. Show that for Mercury the predicted perihelion advance is 43.0 arcsecond/century.

The perihelion of Mercury advances ~ 500 arcsec/century. Since 1859 astronomers explained all but about 43 arcsec as due to slight gravitational effects of other planets. The 43 arcsec was the only known discrepancy between Newton's theory and observation when Einstein published the general theory of relativity. Less than an *arcminute* per *century*! This is the remarkable accuracy of Newton's theory. The explanation of the discrepancy was the first observational verification of general relativity. Today this prediction of general relativity is verified within .1%.

Light Deflection. General relativity predicts that light passing near the Sun will be deflected. See Fig. 3.1. Consider light which grazes the Sun at $\theta = 0$ when $u = u_p = 1/r_p$. Set $E = 0$ in Eq. (3.19) for light to give the geodesic equation

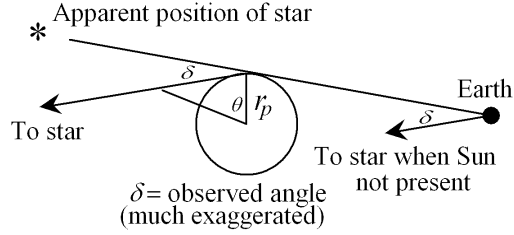


Fig. 3.1: Deflection of light by the sun.

$$\frac{d^2u}{d\theta^2} + u = 3mu^2. \quad (3.24)$$

As a first approximation solve Eq. (3.24) without the small $3mu^2$ term: $u = u_p \cos \theta$. This is the polar equation of a straight line. Substitute $u = u_p \cos \theta$ into the right side of Eq. (3.24) and solve to obtain a better approximation:

$$u = u_p \cos \theta + \frac{1}{2}mu_p^2(3 - \cos 2\theta).$$

Set $u = 0$ ($r = \infty$; $\theta = \pm(\pi/2 + \delta/2)$) and approximate

$$\cos(\pi/2 + \delta/2) = -\sin \delta/2 \approx -\delta/2, \quad \cos(\pi + \delta) = -\cos \delta \approx -1$$

to obtain the observed deflection angle $\delta = 4m/r_p$. (The Newtonian equation Eq. (2.1) predicts half of this deflection.)

Exercise 3.10. Show that for the Sun, $\delta = 1.75$ arcsec.

Stars can be seen near the Sun only during a solar eclipse. Sir Arthur Stanley Eddington organized an expedition to try to detect the deflection during the eclipse of 1919. He confirmed the prediction of general relativity within about 20%. Later eclipse observations have improved the accuracy, but not by much.

In 1995 the predicted deflection was verified within .1% using radio waves emitted by quasars. Radio waves have two advantages. First, radio telescopes thousands of kilometers apart can work together to measure angles more accurately than optical telescopes. Second, an eclipse is not necessary, as radio

sources can be detected during the day. The accuracy of such measurements is now better than 10^{-4} arcsec. Light arriving from a quasar at an angle of 90° from the sun is deflected by 4×10^{-3} arcsec by the sun, and so the deflection can be measured.

A spectacular example of the gravitational deflection of light was discovered in 1979. Two quasars, 6 arcsec apart, are in fact the same quasar! Fig. 3.2 shows the cause of the double image of the quasar. The galaxy is a *gravitational lens*. Many multiple (up to 7) image quasars are now known. It is not necessary that the quasar be exactly centered behind the galaxy for this to occur. But in many cases this is nearly so; then the quasar appears as one or several arcs around the galaxy. And in at least one case the quasar appears as a ring around the galaxy! A gravitational lens can also brighten a distant quasar by up to 100 times, enabling astronomers to study them better.

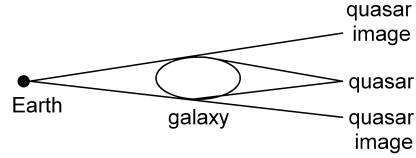


Fig. 3.2: Gravitational lens.

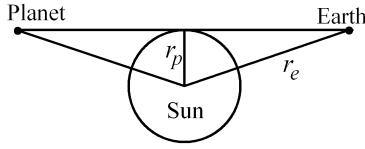


Fig. 3.3: Radar echo delay.

Light Delay. Radar can be sent from Earth, reflected off a planet, and detected upon its return to Earth. If this is done as the planet is about to pass behind the Sun, then according to general relativity, the radar's return is delayed by the Sun's gravity. See Fig. 3.3.

To calculate the time t for light to go from Earth to Sun, use Eq. (3.14):

$$\frac{dr}{dp} = \frac{dr}{dt} \frac{dt}{dp} = \frac{dr}{dt} B \left(1 - \frac{2m}{r} \right)^{-1}.$$

Set $r = r_p$ and $\dot{r}|_{r=r_p} = 0$ in Eq. (3.16):

$$A^2 B^{-2} = r_p^2 \left(1 - \frac{2m}{r_p} \right)^{-1}.$$

Divide Eq. (3.16) by B^2 , substitute the above two equations into the result, separate the variables, and integrate:

$$t = \int_{r_p}^{r_e} \left(\frac{\left(1 - \frac{2m}{r} \right)^{-2}}{1 - \frac{1-2m/r}{1-2m/r_p} \left(\frac{r_p}{r} \right)^2} \right)^{\frac{1}{2}} dr. \quad (3.25)$$

In Appendix 14 we approximate the integral:

$$t = (r_e^2 - r_p^2)^{\frac{1}{2}} + 2m \ln(2r_e/r_p) + m. \quad (3.26)$$

The first term, $(r_e^2 - r_p^2)^{\frac{1}{2}}$, is, by the Pythagorean theorem, the time required for light to travel in a straight line in a flat spacetime from Earth to Sun. The other terms represent a delay of this flat spacetime t .

Add analogous delay terms for the path from Sun to Planet, double to include the return trip, and find for Mercury a delay of 2.4×10^{-4} sec. One difficulty in performing the experiment lies in determining the positions of the planets in terms of the r coordinate of the Schwartzschild metric accurately enough to calculate the flat spacetime time. This difficulty and others have been overcome and the prediction verified within 5%. Signals from the Viking spacecraft on Mars have confirmed this effect within .1%

Quasars can vary in brightness on a timescale of months. We see variation in the southern image of the original double quasar about $1\frac{1}{2}$ years after the same variation of the northern image. The southern image of the quasar is 1 arcsec from the lensing galaxy; that of the northern is 5 arcsec. Thus the light of the southern image travels less distance than that of the southern, causing a delay in the northern image. But this is more than compensated for by the gravitational delay of the light of the southern image passing closer to the lensing galaxy.

Geodetic Precession. Another effect of general relativity, Willem deSitter's *geodetic precession*, is motivated in Figs. 3.4 and 3.5. In Fig. 3.4, a vector in a plane is moved parallel to itself from A around a closed curve made up of geodesics. The vector returns to A with its original direction. In Fig. 3.5 a vector on a sphere is moved parallel to itself (i.e., parallel to itself in local planar frames), from A around a closed curve made of geodesics. The vector returns to A rotated through the angle α . This is a manifestation of curvature.

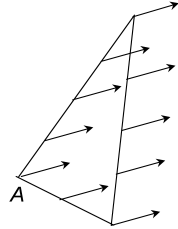


Fig. 3.4: A parallel transported vector returns to A with its original direction.

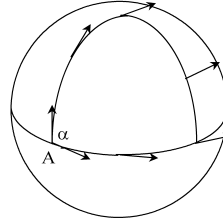


Fig. 3.5: A parallel transported vector returns to A rotated through angle α .

The axis of rotation of an inertial gyroscope moves parallel to itself in inertial frames in a flat spacetime and in local inertial frames in a curved spacetime. But in a curved spacetime the orientation of the axis with respect to the distant stars can change over a worldline. This is geodetic precession. The Earth-Moon system is a "gyroscope" orbiting the Sun. The geodetic precession predicted is a change of the axis of rotation of $\sim .02$ arcsec/yr. The lunar laser experiment confirmed this to 1%.

3.4 Kerr Spacetimes

The Schwarzschild metric describes the spacetime around a spherically symmetric object. In particular, the object cannot be rotating. The *Kerr metric*, discovered only in 1963 by Roy Kerr, describes the spacetime outside a rotating object which is symmetric around its axis of rotation and unchanging in time:

Kerr Metric

$$ds^2 = \left(1 - \frac{2mr}{\Sigma}\right) dt^2 + \frac{4mar \sin^2 \phi}{\Sigma} dt d\theta - \frac{\Sigma}{\Delta} dr^2 - \Sigma d\phi^2 - \left(r^2 + a^2 + \frac{2ma^2 r \sin^2 \phi}{\Sigma}\right) \sin^2 \phi d\theta^2, \quad (3.27)$$

where $m = \kappa M$, as in the Schwarzschild metric; $a = J/M$, the angular momentum per unit mass of the central object; $\Sigma = r^2 + a^2 \cos^2 \phi$; $\Delta = r^2 - 2mr + a^2$; and the axis of rotation is the z -axis. The parameter a is restricted to $0 \leq a \leq 1$.

Exercise 3.11. Show that if the central object is not spinning, then the Kerr metric Eq. (3.27) reduces to the Schwarzschild metric Eq. (3.5).

Note the surfaces $t = \text{const}$, $r = \text{const}$ in a Kerr spacetime do not have the metric of a sphere.

Gravitomagnetism. All of the gravitational effects that we have discussed until now are due to the *mass* of an object. *Gravitomagnetism* is the gravitational effect of the *motion* of an object. Its name refers to an analogy with electromagnetism: an electric charge at rest creates an electric field, while a moving charge also creates a magnetic field.

The *Lense-Thirring* or *frame dragging* gravitomagnetic effect causes a change in the orientation of the axis of rotation of a small spinning object in the spacetime around a large spinning object. An Earth satellite is, by virtue of its orbital motion, a spinning object, whose axis of rotation is perpendicular to its orbital plane. The Kerr metric predicts that the axis of a satellite in polar orbit at two Earth radii will change by $\sim .03$ arcsec/yr. This is a shift of ~ 1 m/yr in the intersection of the orbital and equatorial planes. Observations of two LAGEOS satellites in non-polar orbits have detected this within, it is claimed, 10% of the prediction of general relativity. This is the first direct measurement of gravitomagnetism.

The *Gravity Probe B* experiment, launched in polar orbit in April 2004, will accurately measure frame dragging on gyroscopes. The experiment is a technological tour-de-force. The gyroscopes are rotating ping-pong ball sized spheres made of fused quartz, polished within 40 atoms of a perfect sphere, and cooled to near absolute zero with liquid helium. To work properly, the gyroscopes must move exactly on a geodesic. In particular, atmospheric drag

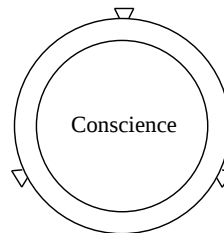


Fig. 3.6: The outer shell and rockets keep the conscience inertial.

and the solar wind must not affect the gyroscopes' orbit. To accomplish this, the satellite is *drag free*: the gyroscopes are in an inner free floating part of the satellite, its *conscience*. The conscience is protected from outside influences by an outer shell. See Fig. 3.6. Small rockets on the shell keep it centered around the conscience when the shell deviates from inertial motion. Clever!

The predicted frame dragging precession of a gyroscope whose axis is perpendicular to its orbit is .04 arcsec/year. The experiment will test this to about 1%. The experiment will also measure geodetic precession. The predicted geodetic precession of a gyroscope whose axis is in the plane of its orbit is 7 arcsec/year. The experiment will test this to better than one part in 10^4 . Results are expected in 2006.

The frame dragging effects of the rotating Earth are tiny. But frame dragging is an important effect in *accretion disks* around a rotating black hole. These disks are a distant relative of the rings of Saturn. They form when a black hole in a binary star system attracts matter from its companion, forming a rapidly rotating disk of hot gas around it. There is evidence that frame dragging causes the axis of some disks to precess hundreds of revolutions a second.

Satellites in the same circular equatorial orbit in a Kerr spacetime but with opposite directions take different coordinate times $\Delta t_{\pm}$ for the complete orbits $\theta \rightarrow \theta \pm 2\pi$. (Note that the orbits do not return the satellite to the same point with respect to the spinning Earth's surface.) We derive this *gravitomagnetic clock effect* from the r -geodesic equation for circular equatorial orbits in the Kerr metric:

$$\dot{t}^2 - 2at\dot{\theta} + \left(a^2 - \frac{r^3}{m}\right)\dot{\theta}^2 = 0. \quad (3.28)$$

Solve for $dt/d\theta$:

$$\left(\frac{dt}{d\theta}\right)_{\pm} = \pm \left(\frac{r^3}{m}\right)^{\frac{1}{2}} + a, \quad (3.29)$$

where $\pm$ correspond to increasing and decreasing θ .

Integrate Eq. (3.29) over one orbit. This multiplies the right side of the equation by $\pm 2\pi$: $\Delta t_{\pm} = 2\pi(r^3/m)^{\frac{1}{2}} \pm 2\pi a$. Then $\Delta t_{+} - \Delta t_{-} = 4\pi a$; an orbit in the sense of a takes longer than an opposite orbit.

Exercise 3.12. Show that for the Earth, $\Delta t_{+} - \Delta t_{-} = 2 \times 10^{-7}$ sec. This has not been measured.

3.5 The Binary Pulsar

We have seen that the differences between Einstein's and Newton's theories of gravity are exceedingly small in the solar system. Large differences require strong gravity and/or large velocities. Both are present in a remarkable binary star system discovered in 1974: two 1.4 solar mass neutron stars have an orbital period of eight hours and are about a solar diameter apart. Imagine!

By great good fortune, one of the neutron stars is a pulsar spinning at 17 revolutions/sec. The pulsar's pulses serve as the ticking of a very accurate clock in the system, rivalling the most accurate clocks on Earth. Several relativistic effects have been measured by analyzing the arrival times of the pulses at Earth. These measurements provide the most probing tests of general relativity to date, earning the system the title "Nature's gift to relativists".

The periastron advance of the pulsar is $4.2^\circ/\text{year}$ – 35,000 times that of Mercury. (Perihelion refers to the Sun; periastron is the generic term.) The Schwarzschild metric cannot predict this because neither star is an inertial particle in the gravity of the other. More complicated techniques using the full field equation, rather than the vacuum field equation, must be used.

Several changing relativistic effects affecting the arrival times at Earth of the pulsar's pulses have been measured: the gravitational redshift of the rate of clocks on Earth as the Earth's distance from the Sun changes over the course of the year, the gravitational redshift of the time between the pulses as the pulsar moves in its highly elliptical (eccentricity = .6) orbit, time dilation of the pulsar clock due to its motion, and light delay in the pulses.

Most important, the system provides evidence for the existence of the gravitational waves predicted by general relativity. Gravitational waves are propagated disturbances in the metric caused by matter in motion. They travel at the speed of light and carry energy away from their source. The mathematical description of gravitational waves and their physical interpretation are complicated matters beyond the scope of this book. Violent astronomical phenomena, such as supernovae (see Sec. 3.1), and the big bang (see Sec. 4.1) emit gravitational waves. Several detectors have been built and are being built to try to detect gravitational waves. So far none has been detected. The difficulty in detecting them is their extreme weakness.

The binary pulsar system provides indirect evidence for the existence of gravitational waves. Calculations predict that the orbital period of the binary pulsar system should decrease by 10^{-4} sec/year due to energy loss from the system due to gravitational waves. By 1978 the period was known to decrease by about the amount predicted by general relativity. The prediction is now confirmed within .5%.

An even more spectacular binary system of neutron stars was discovered in 2003: *both* neutron stars are pulsars. They are 2000 light years away, are about a solar radius apart, and orbit each other in 2.4 hours. One of the pulsars spins at 43 revolutions/sec and the other at .4 revolutions/sec. A few years of observation of this system should greatly improve the accuracy of some existing tests of general relativity and also provide new tests.

3.6 Black Holes

The Schwarzschild metric Eq. (3.5) has a singularity at the *Schwarzschild radius* $r = 2m$: $g_{rr} = \infty$ there. For the Sun, the Schwarzschild radius is 3 km and for the Earth it is .9 cm, well inside both bodies. Since the Schwarzschild metric is a solution of the vacuum field equation, valid only outside the central body, the singularity has no physical relevance for the Sun or Earth.

We may inquire however about the properties of an object which is inside its Schwarzschild radius. The object is then a *black hole*. The $r = 2m$ surface is called the *(event) horizon* of the black hole.

Exercise 3.13. The metric of a plane using polar coordinates is $ds^2 = dr^2 + r^2 d\theta^2$. Substitute $u = 1/r$ to obtain the metric $ds^2 = u^{-4} du^2 + u^{-2} d\theta^2$. This metric has a singularity at the origin, even though there is no singularity in the plane there. It is called a *coordinate singularity*.

Introduce *Painlevé coordinates* with the coordinate change

$$dt = d\bar{t} + \frac{(2mr)^{\frac{1}{2}}}{r - 2m} dr$$

in the Schwarzschild metric, Eq. (3.5), to obtain the metric

$$ds^2 = d\bar{t}^2 - \left(dr + \left(\frac{2m}{r} \right)^{\frac{1}{2}} d\bar{t} \right)^2 - r^2 d\Omega^2.$$

The only singularity in this metric is at $r = 0$. This shows that the singularity at the horizon of the Schwarzschild metric is only a coordinate singularity; there is no singularity in the spacetime. An observer crossing the horizon would not notice anything special. The $r = 0$ singularity is a true spacetime singularity.

Suppose a material particle or pulse of light moves outward (not necessarily radially) from $r_1 = 2m$ to r_2 . Since $ds^2 \geq 0$ and $d\Omega^2 \geq 0$, the Schwarzschild metric Eq. (3.5) shows that $dt \geq (1 - 2m/r)^{-1} dr$. Thus the total coordinate time Δt satisfies

$$\Delta t \geq \int_{2m}^{r_2} \left(1 - \frac{2m}{r} \right)^{-1} dr = \int_{2m}^{r_2} \left(1 + \frac{2m}{r - 2m} \right) dr = \infty;$$

neither matter nor light can escape a Schwarzschild black hole! The same calculation shows that it takes an infinite coordinate time Δt for matter or light to enter a black hole.

Exercise 3.14. Suppose an inertial object falling radially toward a black hole is close to the Schwarzschild radius, so that $r - 2m$ is small. Use Eqs. (3.14) and (3.17) and approximate to show that $dr/dt = (2m - r)/2m$. Integrate and show that $r - 2m$ decreases exponentially in t .

According to Eq. (3.8), the redshift of an object approaching a Schwarzschild black hole increases to infinity as the object approaches the Schwarzschild radius. A distant observer will see the rate of all physical processes on the object slow to zero as it approaches the Schwarzschild radius. In particular, its brightness (rate of emission of light) will dim to zero and it will effectively disappear.

On the other hand, an inertial object radially approaching a Schwarzschild black hole will cross the Schwarzschild radius in a finite time according to a clock it carries! To see this, observe that by the metric postulate Eq. (2.15), dr/ds is the rate of change of r measured by a clock carried by the object. By Eq. (3.17), $|dr/ds|$ increases as r decreases. Thus the object will cross the Schwarzschild radius, never to return, in a finite proper time Δs measured by its clock.

Most stars rotate. Due to conservation of angular momentum, a rotating star collapsed to a black hole will spin very rapidly. We must use the Kerr metric to describe a rotating black hole.

For a Schwarzschild black hole, $g_{rr} = \infty$ and $g_{tt} = 0$ on the surface $r = 2m$, the horizon. We have seen that there is no singularity in the spacetime there. For a Kerr black hole, $g_{rr} = \infty$ on the surface $\Delta = 0$, and $g_{tt} = 0$ on the surface $\Sigma = 2mr$. Fig. 3.7 shows the axis of rotation of the Kerr black hole, the inner surface $\Delta = 0$ (called the *horizon*), and the outer surface $\Sigma = 2mr$ (called the *static limit*). There is no singularity in the spacetime on either surface.

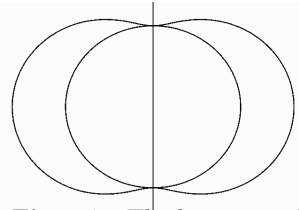


Fig. 3.7: The horizon and static limit of a Kerr black hole.

- Exercise 3.15.** a. Show that the horizon is at $r = m + (m^2 - a^2)^{\frac{1}{2}}$.
 b. Show that the static limit is at $r = m + (m^2 - a^2 \cos^2 \phi)^{\frac{1}{2}}$.

As with the Schwarzschild horizon, it is impossible to escape the Kerr horizon, even though it is possible to reach it in finite proper time. The proofs are more difficult than for the Schwarzschild metric. They are not given here.

The crescent shaped region between the horizon and the static limit is the *ergosphere*. In the ergosphere, the dt^2 , dr^2 , $d\phi^2$, and $d\theta^2$ terms in the Kerr metric Eq. (3.27) are negative. Since $ds^2 \geq 0$ for both particles and light, the $dt d\theta$ term must be positive, i.e., $a(d\theta/dt) > 0$. In the ergosphere it is impossible to be at rest, and motion *must* be in the direction of the rotation!

General relativity allows black holes, but do they exist in Nature? The answer is almost certainly “yes”. At least two varieties are known: *solar mass* black holes, of a few solar masses, and *supermassive* black holes, of $10^6 - 10^{10}$ solar masses.

We saw in Sec. 3.1 that the Earth, the Sun, white dwarf stars, and neutron stars are stabilized by different forces. No known force can stabilize a supernova remnant larger than ~ 3 solar masses; it will collapse to a black hole.

Since black holes emit no radiation, they must be observed indirectly. Several black holes of a few solar masses are known in our galaxy. They are members of X-ray emitting binary star systems in which a normal star is orbiting with a compact companion. Orbital data show that in some of these systems, the companion’s mass is small enough to be a neutron star. But in others, the companion is well over the three solar mass limit for neutron stars, and is therefore believed to be a black hole.

During “normal” periods, gas pulled from the normal star and heated violently on its way to the compact companion causes the X-rays. Some of the systems with a neutron star companion exhibit periods of much stronger X-ray emission. Sometimes the cause is gas suddenly crashing onto the surface of the neutron star. Sometimes the cause is a thermonuclear explosion of material accumulated on the surface. Systems with companions over three solar masses never exhibit such periods. Presumably this is because the companions are black holes, which do not have a surface.

Many, if not most, galaxies have a supermassive black hole at their center. X-ray observations of several galactic nuclei indicate that matter orbiting the nuclei at 3-10 Schwarzschild radii at 10% of the speed of light. Quasars are apparently powered by a massive black hole at the center of a galaxy.

There is a supermassive black hole at the center of our galaxy. There is a star which orbits the hole with eccentricity .87, closest approach to the hole 17 light days, and period 15 years. This implies that the black hole is ≈ 4 million solar masses. The hole is spinning rapidly: there is good evidence that the angular momentum parameter a in the Kerr metric, Eq. (3.27), is about half of the maximum possible value $a = 1$.

Chapter 4

Cosmological Spacetimes

4.1 Our Universe I

This chapter is devoted to *cosmology*, the study of the universe as a whole. Since general relativity is our best theory of space and time, we use it to construct a model of the spacetime of the entire universe. This is an audacious move, but the model is very successful, as we shall see.

We begin with a description of the universe as seen from Earth. The Sun is but one star of 200 billion or so bound together gravitationally to form the *Milky Way galaxy*. It is but one of tens of billions of galaxies in the visible universe. It is about 100,000 light years across. Galaxies cluster in groups of a few to a few thousand. Our Milky Way is a member of the *local group*, consisting of over 30 galaxies. The nearest major galaxies are a few million light years away. Galaxies are inertial objects.

All galaxies, except a few nearby, recede from us with a velocity v proportional to their distance d from us:

$$v = Hd, \quad (4.1)$$

This is the “expansion of the universe”. Eq. (4.1) is *Hubble’s law*; H is *Hubble’s constant*. Its value is 22 (km/sec)/million light years, within 5%.

An expanding spherical balloon, on which bits of paper are glued, each representing a galaxy, provides a very instructive analogy. See Fig. 4.1. The balloon’s two dimensional surface is the analog of our universe’s three dimensional space. From the viewpoint of *every* galaxy on the balloon, other galaxies recede, and Eq. (4.1) holds. The galaxies are glued, not painted on, because they do not expand with the universe.

We see the balloon expanding in an already existing three dimensional space, but surface dwellers are not aware of this third spatial dimension. For both sur-

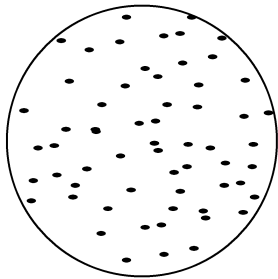


Fig. 4.1: Balloon analogy to the universe.

face dwellers in their universe and we in ours, galaxies separate not because they are moving apart in a static fixed space, but because *space itself is expanding*, carrying the galaxies with it.

We cannot verify Eq. (4.1) directly because neither d nor v can be measured directly. But they can be measured indirectly, to moderate distances, as follows.

The distance d to objects at moderate distances can be determined from their luminosity (energy received at Earth/unit time/unit area). If the object is near enough so that relativistic effects can be ignored, then its luminosity ℓ at distance d is its intrinsic luminosity L (energy emitted/unit time) divided by the area of a sphere of radius d : $\ell = L/4\pi d^2$. Thus if the intrinsic luminosity of an object is known, then its luminosity determines its distance. The intrinsic luminosity of various types of stars and galaxies has been approximately determined, allowing an approximate determination of their distances.

The velocity v of galaxies to moderate distances can be determined from their redshift. Light from all galaxies, except a few nearby, exhibits a redshift $z > 0$. By definition, $z = f_e/f_o - 1$ (see Eq. (1.7)). If f_e is known (for example if the light is a known spectral line), then z is directly and accurately measurable. Eq. (4.11) shows that the redshifts are not Doppler redshifts, but *expansion redshifts*, the result of the expansion of space.

However, to moderate distances we can ignore relativistic effects and determine v using the approximate Doppler formula $z = v$ from Exercise 1.6 b. Substituting $z = v$ into Eq. (4.1) gives $z = Hd$, valid to moderate distances.¹ This *approximation* to Hubble's law was established in 1929 by Edwin Hubble. Since z increases with d , we can use z (which is directly and accurately measurable), as a proxy for d (which is not). Thus we say that a galaxy is “at”, e.g., $z = 2$.

The Big Bang. Reversing time, the universe contracts. From Eq. (4.1) galaxies approach us with a velocity proportional to their distance. Thus they all arrive here at the same time. Continuing into the past, galaxies were crushed together. The matter in the universe was extremely compressed, and thus at an extremely high temperature, accompanied by extremely intense electromagnetic radiation. Matter and space were taking part in an explosion, called the *big bang*. The recession of the galaxies that we see today is the continuation of the explosion. The big bang is the origin of our universe. It occurred ~ 13 billion years ago. The idea of a hot early universe preceding today's expansion was first proposed by Georges Lemaître in 1927.

There is no “site” of the big bang; it occurred everywhere. Our balloon again provides an analogy. Imagine it expanding from a point, its big bang. At a later time, there is no specific place on its surface where the big bang occurred.

Due to the finite speed of light, we see a galaxy not as it is today, but as it was when the light we detect from it was emitted. When we look out in space, we look back in time! This allows us to study the universe at earlier times. Fig. 4.2 illustrates this. Telescopes today see some galaxies as they were when the universe was less than 1 billion years old.

¹The precise relation between z and d is given by the distance-redshift relation, derived in Appendix 16.

The Cosmic Background Radiation. In 1948 Ralph Alpher and Robert Herman predicted that the intense electromagnetic radiation from the big bang, much diluted and cooled by the universe's expansion, still fills the universe. In 1965 Arno Penzias and Robert Wilson, by accident and unaware of the prediction, discovered microwave radiation with the proper characteristics. This *cosmic background radiation* (CBR) provides strong evidence that a big bang really occurred. In Sec. 4.4 we shall see that it contains a wealth of cosmological information.

The CBR has a 2.7 K blackbody spectrum. The radiation started its journey toward us $\sim 380,000$ years after the big bang, when the universe had cooled to about 3000 K, allowing nuclei and electrons to condense into neutral atoms, which made the universe transparent to the radiation.

As a first approximation the CBR is isotropic, i.e., spherically symmetric about us. However, in 1977 astronomers measured a small dipole anisotropy: a relative redshift in the CBR of $z = -.0012 \cos \theta$, where θ is the angle from the constellation Leo.

Exercise 4.1. Explain this as due to the Earth moving toward Leo at 370 km/sec with respect to an isotropic CBR. In this sense we have discovered the absolute motion of the Earth.

The relative redshift is modulated by the orbital motion of the Earth around the Sun. After correcting for the dipole anisotropy, the CBR is isotropic to 1 part in 10,000.

Further evidence for a big bang comes from the theory of *big bang nucleosynthesis*. The theory predicts that from a few seconds to a few minutes after the big bang, nuclei formed, almost all hydrogen (75% by mass) and helium-4 (25%), with traces of some other light elements. Observations confirm the prediction. This is a spectacular success of the big bang theory. Other elements contribute only a small fraction of the mass of the universe. They are mostly formed in stars and are strewn into interstellar space by supernovae and by other means, to be incorporated into new stars, their planets, and any life which may arise on the planets.

The big bang is part of the most remarkable generalization of science: *The universe had an origin and is evolving at many interrelated levels.* For example, galaxies, stars, planets, life, and cultures all evolve.

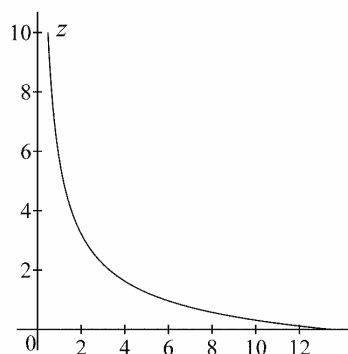


Fig. 4.2: The redshift z of a galaxy vs. the age of the universe (in 10^9 years) when the light we see from it was emitted.

4.2 The Robertson-Walker Metric

The CBR is highly isotropic. Galaxies are distributed approximately isotropically about us. Redshift surveys of hundreds of thousands of galaxies show that our position in the universe is not special. For our cosmological model we assume that the universe is exactly isotropic about every point. This smears the galaxies into a uniform distribution of matter. This should not affect the large scale structure of the universe, which is what we are interested in.

We first set up a coordinate system (t, r, ϕ, θ) for the universe. As the universe expands, the density ρ of matter decreases. Place a clock at rest in every galaxy and set it to some agreed time when some agreed value of ρ is observed at the galaxy. The t coordinate is the (proper) time measured by these clocks.

Choose any galaxy as the spatial origin. Isotropy demands that the ϕ and θ coordinates of other galaxies do not change in time. Choose any time t_0 . Define the r coordinate at t_0 to be any quantity that increases with increasing distance from the origin. Galaxies in different directions but at the same distance are assigned the same r . Now define the r coordinate of galaxies at times other than t_0 by requiring that r , like ϕ and θ , does not change. Isotropy ensures that all galaxies at a given r at a given t are at the same distance from the origin.

The coordinates (r, ϕ, θ) are called *comoving*. Distances between galaxies change in time not because their (r, ϕ, θ) coordinates change, but because their t coordinate changes, which changes the metric.

Our balloon provides an analogy. Recall that $d\Omega^2 = d\phi^2 + \sin^2\phi d\theta^2$ is the metric of a unit sphere. If the radius of the balloon at time t is $S(t)$, then the balloon's metric at time t is

$$ds^2 = S^2(t) d\Omega^2. \quad (4.2)$$

If $S(t)$ increases, then the balloon expands. Distances between the glued on galaxies increase, but their comoving (ϕ, θ) coordinates do not change.

Exercise 4.2. Derive the analog of Eq. (4.1) for the balloon. Let d be the distance between two galaxies on the balloon at time t , and let v be the speed with which they are separating. Show that $v = Hd$, where $H = S'(t)/S(t)$. Hint: Peek ahead at the derivation of Eq. (4.9).

Exercise 4.3. Substitute $\sin\phi = r/(1 + r^2/4)$ in Eq. (4.2). Show that the metric with respect to the comoving coordinates (r, θ) is

$$ds^2 = S^2(t) \frac{dr^2 + r^2 d\theta^2}{(1 + r^2/4)^2}. \quad (4.3)$$

Again, distances between the glued on galaxies increase with $S(t)$, but their comoving (r, θ) coordinates do not change.

The metric of an isotropic universe, after suitably choosing the comoving r coordinate, is:

Robertson-Walker Metric

$$\boxed{\begin{aligned} ds^2 &= dt^2 - S^2(t) \frac{dr^2 + r^2 d\Omega^2}{(1 + k r^2/4)^2} \quad (k = 0, \pm 1) \\ &\equiv dt^2 - S^2(t) d\sigma^2. \end{aligned}} \quad (4.4)$$

The elementary but somewhat involved derivation is given in Appendix 15. The field equation is not used. We will use it in Sec. 4.4 to determine $S(t)$. The constant k is the curvature of the spatial metric $d\sigma$: $k = -1, 0, 1$ for negatively curved, flat, and positively curved respectively.

The metric was discovered independently by H. P. Robertson and E. W. Walker in the mid 1930's. It should be compared with the metric in Eq. (4.3).

Distance. Emit a light pulse from a galaxy to a neighboring galaxy. Since $ds = 0$ for light, Eq. (4.4) gives $dt = S(t) d\sigma$. By the definition of t , or from Eq. (4.4), t is measured by clocks at rest in galaxies. Thus dt is the elapsed time in a local inertial frame in which the emitting galaxy is at rest. Since $c = 1$ in local inertial frames, the distance between the galaxies is

$$S(t) d\sigma. \quad (4.5)$$

This is the distance that would be measured by a rigid rod. The distance is the product of two factors: $d\sigma$, which does not change in time, and $S(t)$, which is independent of position. Thus S is a scale factor for the universe: if, e.g., $S(t)$ doubles over a period of time, then so does the distance between all galaxies.

The metric of points at coordinate radius r in the Robertson-Walker metric Eq. (4.4) is that of a sphere of radius $S(t) r / (1 + k r^2/4)$. From Eq. (4.5) this metric measures physical distances. Thus the sphere has surface area

$$A(r) = \frac{4\pi r^2 S^2(t)}{(1 + k r^2/4)^2} \quad (4.6)$$

and volume

$$V(r) = S(t) \int_0^r \frac{A(\rho)}{1 + k \rho^2/4} d\rho. \quad (4.7)$$

If $k = -1$ (negative spatial curvature), then r is restricted to $0 \leq r < 2$. The substitution $\rho = 2 \tanh \xi$ shows that $V(2) = \infty$, i.e., the universe has infinite volume. If $k = 0$ (zero spatial curvature), $0 \leq r < \infty$ and the universe has infinite volume. If $k = 1$ (positive spatial curvature), then $A(r)$ increases as r increases from 0 to 2, but then decreases to zero as $r \rightarrow \infty$. The substitution $\rho = 2 \tan \xi$ shows that $V(\infty) = 2\pi^2 S^3(t)$; the universe has finite volume! The surface of our balloon provides an analogy. It is finite and has positive curvature. Circumferences of circles of increasing radius centered at the North pole increase until the equator is reached and then decrease to zero.

4.3 The Expansion Redshift

At time t the distance from Earth at $r = 0$ to a galaxy at $r = r_e$ is, by Eq. (4.5),

$$d = S(t) \int_0^{r_e} d\sigma. \quad (4.8)$$

This is the distance that would be obtained by adding the lengths of small rigid rods laid end to end between Earth and the galaxy at time t .

Differentiate to give the velocity of the galaxy with respect to the Earth:

$$v = S'(t) \int_0^{r_e} d\sigma.$$

Divide to give Hubble's law $v = Hd$, Eq. (4.1), where

$$\boxed{H = \frac{S'(t)}{S(t)}}. \quad (4.9)$$

From $v = Hd$ we see that if $d = 1/H \approx 13.6$ billion light years, then $v = 1$, the speed of light. Galaxies beyond this distance are today receding from us faster than the speed of light. What about the rule “nothing can move faster than light”? In special relativity the rule applies to objects moving in an inertial frame, and in general relativity to objects moving in a local inertial frame. There is no local inertial frame containing both the Earth and such galaxies.

The distance $d = 1/H$ is at $z = 1.5$. This follows from the distance-redshift relation, derived in Appendix 16.

We now obtain a relationship between the expansion redshift of light from a galaxy and the size of the universe when the light was emitted. Suppose that light signals are emitted toward us at events (t_e, r_e) and $(t_e + \Delta t_e, r_e)$ and received by us at events $(t_o, 0)$ and $(t_o + \Delta t_o, 0)$. Since $ds = 0$ for light, Eq. (4.4) gives

$$\frac{dt}{S(t)} = \frac{dr}{1 + k r^2/4}. \quad (4.10)$$

Integrate this over both light worldlines to give

$$\int_{t_e}^{t_o} \frac{dt}{S(t)} = \int_0^{r_e} \frac{dr}{1 + k r^2/4} = \int_{t_e + \Delta t_e}^{t_o + \Delta t_o} \frac{dt}{S(t)}.$$

Subtract $\int_{t_e + \Delta t_e}^{t_o}$ from both ends to give

$$\int_{t_e}^{t_e + \Delta t_e} \frac{dt}{S(t)} = \int_{t_o}^{t_o + \Delta t_o} \frac{dt}{S(t)}.$$

If Δt_e and Δt_o are small, this becomes

$$\frac{\Delta t_e}{S(t_e)} = \frac{\Delta t_o}{S(t_o)}.$$

Use Eq. (1.6) to obtain the desired relation:

$$\boxed{z + 1 = \frac{S(t_o)}{S(t_e)}}. \quad (4.11)$$

This wonderfully simple formula directly relates the expansion redshift to the universe's expansion. Since $S = 0$ at the big bang, the big bang is at $z = \infty$.

A galaxy at redshift $z \approx 7$ is the most distant known. The light we see from the galaxy was emitted ~ 750 million years after the big bang. The faint galaxy is magnified at least 25 times by a foreground gravitational lensing galaxy.

According to Eq. (4.11) the galaxy is 8 times farther away from us today than when it emitted the light we detect from it! If we could observe a physical process on the galaxy from Earth, then by Eq. (1.6) it would appear to proceed at $1/8$ of the rate of the same process on Earth.

Distant supernovae exhibit this effect. Type Ia supernovae all have about the same duration. But supernovae out to $z = .8$ (the farthest to which measurements have been made) last about $1 + z$ times longer as seen from Earth.

Suppose that an object at distance z emits blackbody radiation at temperature $T(z)$. According to Planck's law, the object radiates at frequency f_e with intensity proportional to

$$\frac{f_e^3}{e^{hf_e/kT(z)} - 1},$$

where h is Planck's constant and k is Boltzmann's constant. Radiation emitted at f_e is received by us redshifted to f_o , where from Eq. (1.7), $f_e = (z + 1)f_o$. Thus the intensity of the received radiation is proportional to

$$\frac{f_o^3}{e^{hf_o/kT(0)} - 1},$$

where $T(0)$ is defined by

$$z + 1 = \frac{T(z)}{T(0)}. \quad (4.12)$$

The observed spectrum is also blackbody, at temperature $T(0)$.

As stated in Sec. 4.1, the CBR was emitted at ≈ 3000 K and is observed today at 2.7 K. Thus from Eq. (4.12), its redshift $z \approx 3000 \text{ K}/2.7 \text{ K} \approx 1100$.

Analysis of spectral lines in a quasar at $z = 3$ shows that they were excited by an ~ 11 K CBR. Eq. (4.12) predicts this: $T(3) = (3 + 1)T(0) \approx 11$ K. From Eq. (4.11), the universe was then about $1/(1 + 3) = 25\%$ of its present size.

4.4 Our Universe II

The Robertson-Walker metric Eq. (4.4) is a consequence of isotropy alone; the field equation was not used in its derivation. There are two unknowns in the metric: the curvature k and the expansion factor $S(t)$. In this section we first review the compelling evidence that in our universe $k = 0$. This spatial flatness, together with observational data, will force us to modify the field equation. We then solve the modified field equation to determine $S(t)$.

In 1981 Alan Guth proposed that there was a period of exponential expansion, called *inflation*, in the very early universe. The inflation started perhaps 10^{-34} sec after the big bang, lasted perhaps 10^{-35} sec, and expanded the universe by a factor of at least 10^{30} ! As fantastic as inflation seems, it explained several puzzling cosmological observations and steadily gained support over the years. By 1990, most cosmologists believed that inflation occurred. Since then more evidence for inflation has accumulated. Inflation predicts a flat universe.

There is more direct evidence for a flat universe from the CBR. When the CBR was emitted, matter was almost, but not quite, uniformly distributed. The variations originated as random quantum fluctuations. We can “see” them because regions of higher density were hotter, leading to tiny temperature anisotropies ($\sim .06$ K, a few parts in 10^5) in the CBR. The size of these regions does not depend on the curvature of space, but their observed angular size does. In a flat universe, the average temperature difference as a function of angular separation peaks at .8 arcdegree. Observations in 2000 found the peak at .8 arcdegree, providing compelling evidence for a flat universe.

To understand how curvature can affect angular size, consider the everyday experience that the angular size of an object decreases with distance. The rate of decrease is different for differently curved spaces. Imagine surface dwellers whose universe is the surface of the sphere in Fig. 4.3 and suppose that light travels along the great circles of the sphere. According to Exercise 2.9 these circles are geodesics. For an observer at the north pole, the angular size of an object of length D decreases with distance, but at a slower rate than on a flat surface. And when the object reaches the equator, its angular size begins to increase. The object has the same angular size α at the two positions in the figure. On a pseudosphere the angular size decreases at a faster rate than on a flat surface. The relationship between the angular size and redshift of distant objects in our universe is given by the angular size-redshift relation, derived in Appendix 16.

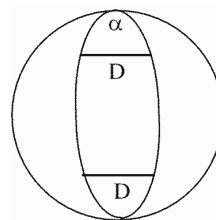


Fig. 4.3: Angular size on a sphere.

The Field Equation. Despite the strong evidence for a flat universe, we shall see that the field equation Eq. (2.27) applied to the Robertson-Walker metric Eq. (4.4) with $k = 0$ gives results in conflict with observation. Several solutions to this problem have been proposed. The simplest is to add a term

$\Lambda \mathbf{g}$ to the field equation:

$$\boxed{\mathbf{G} + \Lambda \mathbf{g} = -8\pi\kappa \mathbf{T}}, \quad (4.13)$$

where Λ is a constant, called the *cosmological constant*. This is the most natural change possible to the field equation. In fact, we have seen the term $\Lambda \mathbf{g}$ before, in the field equation Eq. (2.34). We eliminated it there by assuming that spacetime is flat in the absence of matter. Thus the new term represents a gravitational effect of empty space. If Λ is small enough, then its effects in the solar system are negligible, but it can dramatically affect cosmological models.

To apply the new field equation, we need the energy-momentum tensor $\mathbf{T}$. In Sec. 4.2 we distributed all matter in the universe into a uniform distribution of density $\rho = \rho(t)$. Since galaxies interact only gravitationally, it is reasonable to assume that the matter is dust. (This is unreasonable in the very early universe.) Thus $\mathbf{T}$ is of the form Eq. (2.26). For the comoving matter, $dr = d\theta = d\phi = 0$, and from Eq. (4.4), $dt/ds = 1$. Thus $T^{tt} = \rho$ is the only nonzero T^{jk} .

The field equation Eq. (4.13) for the Robertson-Walker metric Eq. (4.4) with $k = 0$ and the $\mathbf{T}$ just obtained reduces to two ordinary differential equations:

$$-3 \left(\frac{S'}{S} \right)^2 + \Lambda = -8\pi\kappa\rho, \quad (tt \text{ component}) \quad (4.14)$$

$$S'^2 + 2SS'' - \Lambda S^2 = 0. \quad (rr, \phi\phi, \theta\theta \text{ components}) \quad (4.15)$$

From Eqs. (4.14) and (4.9),

$$\boxed{\rho + \frac{\Lambda}{8\pi\kappa} = \rho_c}, \quad (4.16)$$

where

$$\rho_c = 3H^2/8\pi\kappa \quad (4.17)$$

is called the *critical density*. Today $\rho_c \approx 10^{-29} \text{g/cm}^3$ – the equivalent of a few hydrogen atoms per cubic meter.

The field equation implies that the matter density ρ and $\Lambda/8\pi\kappa$ on the left side of Eq. (4.16) add to ρ_c , but their individual contributions in our universe must be determined by observation.

Matter. The best determination of ρ comes from precise measurements of the anisotropies in the CBR made by the *Wilkinson Microwave Anisotropy Probe* (WMAP), and announced in 2003. The clustering of galaxies is sensitive to ρ . Measurements of the positions and redshifts of 200,000 galaxies by the Sloan Digital Sky Survey (SDSS) were also announced in 2003. The SDSS results are entirely consistent with those of WMAP.

According to WMAP the density of matter today is $\rho = .3\rho_c$. This confirmed earlier estimates based on gravitational lens statistics. WMAP measurements

show that some of this matter is in known forms (protons, neutrons, electrons, neutrinos, photons, ...), but most is not.

Known forms of matter contribute only $.05\rho_c$. This confirmed two earlier determinations. One used spectra of distant galactic nuclei to “directly” measure the density of known forms of matter in the early universe. The other inferred the density from the abundance of deuterium created during big bang nucleosynthesis. The deuterium abundance was measured in distant pristine gas clouds backlit by even more distant quasars. It is gratifying to have excellent agreement between the CBR and deuterium measurements, as the deuterium was created a few minutes after the big bang and the CBR was emitted hundreds of thousands of years later. Today most ($\sim 85\%$) of this matter is superhot X-ray emitting intergalactic gas in large galaxy clusters. This leaves only a small fraction in stars and gas in galaxies.

Unknown forms of matter contribute $.25\rho_c$. This *dark matter* does not emit or absorb electromagnetic radiation, and so is not luminous. Astronomers agree that dark matter exists and is much more abundant than ordinary matter, but no one knows what it is.

Gravitationally, dark matter behaves just as ordinary matter. The existence and approximate abundance of dark matter was first inferred from its gravitational effects. For example, stars orbiting in the outer part of a galaxy move too rapidly for the density of ordinary matter to hold them in orbit. Gravitational lensing is used to map the distribution of dark matter. In one system, a quasar is lensed by a foreground cluster of galaxies. The separation of the quasar images is so large that dark matter in the cluster must be responsible for the lensing.

Galaxies would not have formed without dark matter. They arose from the small density variations seen in the CBR. The variations are just the right size to produce the distribution of galaxies that we see today. They did not start with ordinary matter; any such variations were smoothed out by interactions with the intense electromagnetic radiation in the very early universe. But because dark matter does not interact electromagnetically, density variations within it could survive and grow by gravitational attraction.

Dark Energy. From Eq. (4.16) and $\rho = .3\rho_c$, $\Lambda/8\pi\kappa = \rho_c - .3\rho_c = .7\rho_c$. The quantity $\Lambda/8\pi\kappa$ is the density of another mysterious component of our universe, called *dark energy*. Dark energy is intrinsic to space; it literally “comes with the territory”.

Exercise 4.4. Show that $\Lambda \approx 1.2 \times 10^{-35}/\text{sec}^2$.

It would be hasty to accept that $\Lambda/8\pi\kappa = .7\rho_c$, i.e., that 70% of the mass-energy of the universe is in some totally unknown form, on the basis of the above “it is what is needed to satisfy Eq. (4.16)” argument. Fortunately, there is more direct evidence for dark energy from supernovae.

All type Ia supernovae have about the same peak intrinsic luminosity L . The *luminosity-redshift relation*, derived in Appendix 16, expresses the observed luminosity ℓ of an object as a function of its intrinsic luminosity and redshift. The function also depends on Λ . Fig. 4.4 graphs ℓ vs. z for a fixed L and three

values of Λ . The points (z, ℓ) for supernovae in the range $z \approx .5$ to $z \approx 1.5$ lie closest to the $\Lambda/8\pi\kappa = .7\rho_c$ curve, and definitely not the $\Lambda = 0$ curve. This is strong evidence that $\Lambda \neq 0$.

Independent evidence for the existence and density of dark energy comes gravitational lens statistics and from a phenomenon called the *integrated Sachs-Wolfe effect*.

We return to the field equation. A quick calculation using Eq. (4.15) shows that $(SS'^2 - \Lambda S^3/3)' = 0$. Thus from Eq. (4.14),

$$\rho S^3 = \text{constant}. \quad (4.18)$$

According to Eq. (4.18), ρ is inversely proportional to the cube of the scale factor S . This is what we would expect. In contrast, the density of dark energy, $\Lambda/8\pi\kappa$, is constant in time (and space). Thus the ratio of the dark energy density to the matter density increases as the universe expands. At early times matter dominates and at later times dark energy dominates.

This has an important consequence for the expansion of the universe. Solve Eq. (4.14) for S'^2 and substitute into Eq. (4.15), giving $S'' = S(\Lambda - 4\pi\kappa\rho)/3$. At early times ρ is large and so $S'' < 0$; the expansion decelerates. This is no surprise: matter is gravitationally attractive. At later times the dark energy term dominates, and $S'' > 0$; the expansion accelerates! This *is* a surprise: dark energy is gravitationally repulsive.

Exercise 4.5. Show that the light we receive from galaxies at $z \approx .7$ was emitted when the universe made the transition from decelerating to accelerating.

The effect of Λ on the expansion is dramatic: $z \approx .5$ supernovae are $\sim 30\%$ dimmer they would be if Λ were 0. Fig. 4.4 shows this. They are dimmer because their light has farther to travel to get to us, as we accelerate away from it. Supernovae at $z > 1$ show the opposite effect: they are brighter than they would be if the expansion had always been accelerating.

S(t). We now solve the field equation for $S(t)$. Multiply Eq. (4.14) by $-S^3$ and use Eq. (4.18):

$$3SS'^2 - \Lambda S^3 = 8\pi\kappa\rho_o S^3(t_o), \quad (4.19)$$

where a “o” subscript means that the quantity is evaluated today. Integrate:

$$S^3(t) = \frac{8\pi\kappa\rho_o}{\Lambda} \sinh^2\left(\frac{\sqrt{3\Lambda}}{2}t\right) S^3(t_o). \quad (4.20)$$

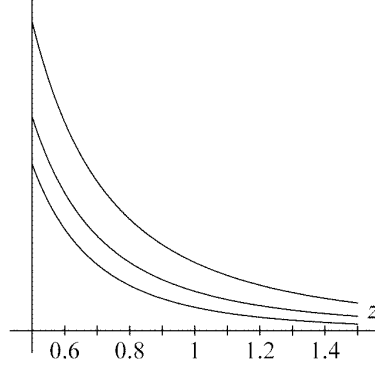


Fig. 4.4: The luminosity-redshift relation. Upper: $\Lambda = 0$; middle: $\Lambda/8\pi\kappa = .7\rho_c$; lower: $\Lambda/8\pi\kappa = \rho_c$.

Since the big bang is the origin of the universe, it is convenient to set $t = 0$ when $S = 0$. This convention eliminates a constant of integration in S . The graph of $S(t)$ is given in Fig. 4.5. As expected, the universe decelerates at early times and then begins to accelerate.

Exercise 4.6. Show that Eq. (4.20) implies that the age of the universe is $t_o = 13.1$ billion years.

Exercise 4.7. Show that Eq. (4.20) implies that in the future the universe will undergo exponential expansion.

Fig. 4.2 graphed the redshift z of a galaxy against the time t it emitted the light we see from it. This relationship between z and t is obtained by substituting Eq. (4.11) into Eq. (4.20).

Summary. The universe had a big bang origin 13 billion years ago. It is flat, expanding, and the expansion is speeding up. It consists of 5% known matter, 25% dark matter, and 70% dark energy. Each of these numbers is supported by independent lines of evidence.

The construction of a cosmological model which fits the observations so well is a magnificent achievement. But the model crystallized only at the turn of the millennium. And we are left with two mysteries: the nature of dark matter and dark energy. They constitute 95% of the “stuff” of the universe. We are in “a golden age of cosmology” in which new cosmological data are pouring in. We must be prepared for changes, even radical changes, in our model.

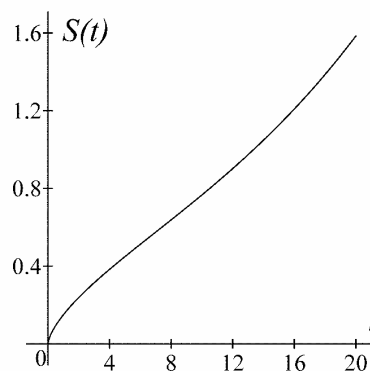


Fig. 4.5: S (in units of $S(t_o)$) as a function of t (in 10^9 years).

Horizons. Light emitted at the big bang has only travelled a finite distance since that time. Our *particle horizon* is the sphere from which we receive today light emitted at the big bang. Galaxies beyond the particle horizon cannot be seen by us today because their light has not had time to reach us since the big bang! Conversely, galaxies beyond the particle horizon cannot see us today. (Actually we can only see most of the way to the horizon, to the CBR.)

From Eqs. (4.8) and (4.4) with $ds = 0$ and $k = 0$, the distance to the particle horizon today is

$$S(t_o) \int_{t=0}^{t_o} \frac{dt}{S(t)}.$$

Exercise 4.8. Show that with $S(t)$ from Eq. (4.20) this distance is ≈ 45 billion light years. Actually, the distance to the horizon highly uncertain, as it is sensitive to the duration and amount of inflation, which Eq. (4.20) does not take into account.

The distance *today* to the particle horizon of another time t_1 is $S(t_o) \int_{t=0}^{t_1} dt/S(t)$. Since the integral increases with t_1 , we can see, and can be seen by, more galaxies with passing time. However, the integral converges as $t_1 \rightarrow \infty$. Thus there are galaxies which will remain beyond the particle horizon forever. The distance today to the most distant galaxies that we will ever see is $S(t_o) \int_{t=0}^{\infty} dt/S(t) \approx 61$ billion light years.

Light emitted today will travel only a finite coordinate distance as $t \rightarrow \infty$. Our *event horizon* is the sphere from which light emitted today reaches us as $t \rightarrow \infty$. We will never see galaxies beyond the event horizon as they are today because light emitted by them today will never reach us!

Exercise 4.9. Show that the distance to the event horizon today is $S(t_o) \int_{t=t_o}^{\infty} dt/S(t)$. This is ≈ 16 billion light years.

The table shows various redshifts z and distances d (in 10^9 lightyears) discussed above. They are related by the distance-redshift relation, derived in Appendix 16. The $S'' = 0$ column is the decelerating/accelerating transition. The $v = c$ column is where galaxies recede at the speed of light.

	$S'' = 0$	$v = c$	Event horizon	CMB	Particle horizon	Most distant ever see
z	.7	1.5	1.8	1100	∞	–
d	7.8	13.6	16	44	45	61

4.5 General Relativity Today

General relativity extended the revolution in our ideas about space and time begun by special relativity. Special relativity showed that space and time are not absolute and independent, but related parts of a whole: spacetime. General relativity showed that a spacetime containing matter is not flat and static, but curved and dynamic, not only affecting matter (according to the geodesic postulate), but also affected by matter (according to the field equation). In consequence, Euclidean geometry, thought inviolate for over twenty centuries, does not apply (exactly) in a gravitational field. The Robertson–Walker spacetimes illustrate these points particularly clearly.

Despite its fundamental nature, general relativity was for half a century after its discovery in 1915 out of the mainstream of physical theory because its results were needed only for the explanation of the few minute effects in the solar system described in Sec. 3.3. Newton’s theory of gravity accounted for all other gravitational phenomena. But since 1960 astronomers have made several spectacular discoveries: quasars, the CBR, neutron stars, the binary pulsar, gravitational lenses, the emission of gravitational radiation, black holes, detailed structure in the CBR, and that the expansion of the universe is speeding up. These discoveries have revealed a wonderfully varied, complex, and interesting universe. They have enormously widened our “cosmic consciousness”. And we need general relativity to help us understand them.

General relativity has passed every test to which it has been put. General relativity reduces to special relativity, an abundantly verified theory, in the absence of significant gravitational fields. General relativity reduces to Newton’s theory of gravity, an exceedingly accurate theory, in weak gravity in which the velocity of matter is small. Sec. 2.2 cited experimental evidence and logical coherence as reasons for accepting the postulates of the theory. The theory explains the minute details of the motion of matter and light in the solar system; the tests of the *vacuum* field equation described in Sec. 3.3 – the perihelion advance of Mercury, light deflection, and light delay – all confirm the predictions of general relativity with an impressive accuracy. The lunar laser experiment confirms tiny general relativistic effects on the motion of the Earth and moon. Gravitomagnetism has been detected. The periastron advance of the binary pulsar and the apparent emission of gravitational radiation from its system provide quantitative confirmation of the *full* field equation. Finally, the application of general relativity on the grandest possible scale, to the universe as a whole, appears to be successful.

As satisfying as all this is, we would like more evidence for such a fundamental theory. This would give us more confidence in applying it. The difficulty in finding tests for general relativity is that the gravitational interaction is weak (about 10^{-40} as strong as the electromagnetic interaction), and so astronomical sized masses over which an experimenter has no control, must be used. Also, Newton’s theory is already very accurate in most circumstances. Rapidly advancing technology has made, and will continue to make, more tests possible.

Foremost among these are attempts to detect gravitational waves. The hope is that they will someday complement electromagnetic waves in providing information about the universe. Sec. 3.5 described observations of the binary pulsar which show indirectly that gravitational waves exist. Several ground based detectors are in operation, and several are planned in space, to try to detect them directly. We still await the first such detection.

The most unsatisfactory feature of general relativity is its conception of matter. The density ρ in the energy-momentum tensor $\mathbf{T}$ represents a continuous distribution of matter, entirely ignoring its atomic and subatomic structure. *Quantum theory* describes this structure. The *standard model* is a quantum theory of subatomic particles and forces. The theory, forged in the 1960's and 1970's, describes all known forms of matter, and the forces between them, *except gravity*. The fundamental particles of the theory are quarks, leptons, and force carriers. There are several kinds of each. Protons and neutrons consist of three quarks. Electrons and neutrinos are leptons. Photons carry the electromagnetic force.

The standard model has been spectacularly successful, correctly predicting the results of all particle accelerator experiments (although many expect discrepancies to appear soon in new experiments). The theory of big bang nucleosynthesis uses the standard model. However, in cosmology the standard model leaves many matters unexplained. For example, inflation, dark matter, and dark energy are all beyond the scope of the theory.

Quantum theory and general relativity are the two fundamental theories of contemporary physics. But they are separate theories; they have not been unified. Thus they must be combined on an ad hoc basis when both gravity and quantum effects are important. For example, in 1974 Steven Hawking used quantum theoretical arguments to show that a black hole emits subatomic particles and light. A black hole is not black! The effect is negligible for a black hole with the mass of the Sun, but it would be important for very small black holes.

The unification of quantum theory and general relativity is the most important goal of theoretical physics today. Despite decades of intense effort, the goal has not been reached. *String theory* and *loop quantum gravity* are candidates for a theory of quantum gravity, but they are very much works in progress.

Despite the limited experimental evidence for general relativity and the gulf between it and quantum theory, it is a necessary, formidable, and important tool in our quest for a deeper understanding of the universe in which we live. This and the striking beauty and simplicity – both conceptually and mathematically – of the unification of space, time, and gravity in the theory make general relativity one of the finest creations of the human mind.

Appendices

A.1 Physical Constants. When appropriate, the first number given uses (light) seconds as the unit of distance.

$$c = 1 = 3.00 \times 10^{10} \text{ cm/sec} = 186,000 \text{ mi/sec (Light speed)}$$

$$\kappa = 2.47 \times 10^{-39} \text{ sec/g} = 6.67 \times 10^{-8} \text{ cm}^3/(\text{g sec}^2)$$

(Newtonian gravitational constant)

$$g = 3.27 \times 10^{-8}/\text{sec} = 981 \text{ cm/sec}^2 = 32.2 \text{ ft/sec}^2 \text{ (Acceleration Earth gravity)}$$

$$\text{Mass Sun} = 1.99 \times 10^{33} \text{ g}$$

$$\text{Radius Sun} = 2.32 \text{ sec} = 6.96 \times 10^{10} \text{ cm} = 432,000 \text{ mi}$$

$$\text{Mean radius Mercury orbit} = 1.9 \times 10^2 \text{ sec} = 5.8 \times 10^{12} \text{ cm} = 36,000,000 \text{ mi}$$

$$\text{Mercury perihelion distance} = 1.53 \times 10^2 \text{ sec} = 4.59 \times 10^{12} \text{ cm}$$

$$\text{Eccentricity Mercury orbit} = .206$$

$$\text{Period Mercury} = 7.60 \times 10^6 \text{ sec} = 88.0 \text{ day}$$

$$\text{Mean radius Earth orbit} = 5 \times 10^2 \text{ sec} = 1.5 \times 10^{13} \text{ cm} = 93,000,000 \text{ mi}$$

$$\text{Mass Earth} = 5.97 \times 10^{27} \text{ g}$$

$$\text{Radius Earth} = 2.13 \times 10^{-2} \text{ sec} = 6.38 \times 10^8 \text{ cm} = 4000 \text{ mi}$$

$$\text{Angular momentum Earth} = 10^{20} \text{ g sec} = 10^{41} \text{ g cm}^2/\text{sec}$$

$$H_o = 2.3 \times 10^{-18}/\text{sec} = (22 \text{ km/sec})/10^6 \text{ light year (Hubble constant)}$$

A.2 Approximations. The approximations are valid for small x .

$$(1+x)^n \approx 1+nx \quad \left[\text{e.g., } (1+x)^{\frac{1}{2}} \approx 1+x/2 \text{ and } (1+x)^{-1} \approx 1-x \right]$$

$$\sin(x) \approx x - x^3/6$$

$$\cos(x) \approx 1$$

A.3 The Macek-Davis Experiment. In this experiment, light was reflected in both directions around a ring laser, setting up a standing wave. A fraction of the light in both directions was extracted by means of a half silvered mirror and their frequencies compared. See Fig. A.6.

Let t_1 and t_2 be the times it takes light to go around in the two directions and f_1 and f_2 be the corresponding frequencies. Then $t_1 f_1 = t_2 f_2$, as both are equal to the number of wavelengths in the standing wave. Thus any fractional difference between t_1 and t_2 would be accompanied by an equal fractional difference between f_1 and f_2 . The frequencies differed by no more than one part in 10^{12} .

(If the apparatus is rotating, say, clockwise, then the clockwise beam will have a longer path, having to “catch up” to the moving mirrors, and the frequencies of the beams will differ. The experiment was performed to detect this effect.)

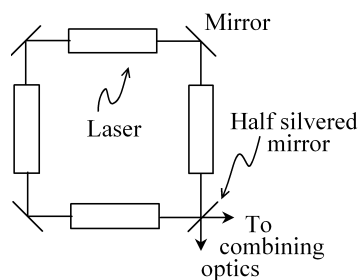


Fig. A.6: Comparing the round trip speed of light in opposite directions.

A.4 Spacelike Separated Events. By definition, spacelike separated events E and F can be simultaneously at the ends of an inertial rigid rod – simultaneously in the sense that light flashes emitted at E and F reach the center of the rod simultaneously, or equivalently, E and F are simultaneous in the rest frame I' of the rod. Since the events are simultaneous in I' , they are neither lightlike separated ($|\Delta x| = |\Delta t|$ in I) nor timelike separated ($|\Delta x| < |\Delta t|$ in I). Thus $|\Delta x| > |\Delta t|$, i.e., $|\Delta t/\Delta x| < 1$.

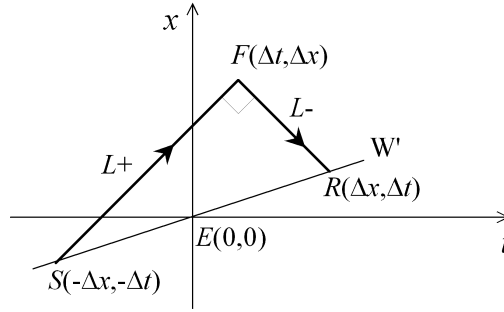


Fig. A.7: $\Delta s^2 = \Delta t^2 - \Delta x^2$ for spacelike separated events E and F . See the text.

For convenience, let E have coordinates $E(0,0)$. Fig. A.7 shows the worldline W' of an inertial observer O' moving with velocity $v = \Delta t/\Delta x$ in I . (That is not a typo.) $L\pm$ are the light worldlines through F . Since $c = 1$ in I , the slope of $L\pm$ is ± 1 . Solve simultaneously the equations for $L-$ and W' to obtain the coordinates $R(\Delta x, \Delta t)$. Similarly, the equations for $L+$ and W' give $S(-\Delta x, -\Delta t)$.

According to Eq. (1.11), the proper time, as measured by O' , between the timelike separated events S and E , and between E and R , is $(\Delta x^2 - \Delta t^2)^{1/2}$. Since the times are equal, E and F are, by definition, simultaneous in an inertial frame I' in which O' is at rest.

Since $c = 1$ in I' , the distance between E and F in I' is $(\Delta x^2 - \Delta t^2)^{1/2}$. This is the length of a rod at rest in I' with its ends simultaneously at E and F . By definition, this is the proper distance $|\Delta s|$ between the events. This proves Eq. (1.11) for spacelike separated events.

A.5 Moving Rods. If an inertial rod is pointing perpendicular to its direction of motion in an inertial frame, then its length, as measured in the inertial frame, is unchanged. If the rod is pointing parallel to its direction of motion in the inertial frame, then its length, as measured in the inertial frame, contracts. Why this difference? An answer can be given in terms of Einstein's principle of relativity: identical experiments performed in different inertial frames give identical results. We show that the principle demands that the length be unchanged in the perpendicular case, but not in the parallel case.

Consider identical rods R and R' , both inertial and in motion with respect to each other. According to the inertial frame postulate, R and R' are at rest in inertial frames, say I and I' .

Perpendicular case. First, situate the rods as in Fig. A.8. Suppose also that the ends A and A' coincide when the rods cross. Then we can compare the lengths of the rods directly, independently of any inertial frame, by observing the relative positions of the ends B and B' when the rods cross. According to the principle of relativity, R and R' have the same length. For if, say, B' passed below B , then this would violate the principle: a rod moving in I is shorter, while an identical rod moving in I' is longer.

Fig. A.8: Rod perpendicular to its direction of motion in I .

Parallel case. Now situate the rods as in Fig. A.9. Wait until A and A' coincide. Consider the relative positions of B and B' at the same time. But "at the same time" is different in I and I' ! What happens is this: At the same time in I , B' will not have reached B , so R' is shorter than R in I . And at the same time in I' , B will have passed B' , so R is shorter than R' in I' . This is *length contraction*. There is no violation of the principle of relativity here: in each frame the moving rod is shorter.

There is no contradiction here: The lengths are compared differently in the two inertial frames, each frame using the synchronized clocks at rest in that frame. Thus there is no reason for the two comparisons to agree. This is different from Fig. A.8, where the lengths are compared directly. Then there can be no disagreement over the relative lengths of the rods.

We calculate the length contraction of R' in I . Let the length of R' in I and I' be D and D' , respectively. Let v be the speed of R' in I . Emit a pulse of light from A' toward B' . In I , B' is at distance D from A' when the light is emitted.

Since $c = 1$ in I , the relative speed of the light and B' in I is $1 - v$. Thus in I the light will take time $D/(1 - v)$ to reach A' . Similarly, if the pulse is reflected at B' back to A' , it will take time $D/(1 + v)$ to reach A' . Thus the

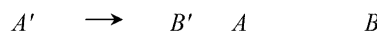


Fig. A.9: Rod parallel to its direction of motion in I .

round trip time in I is

$$\Delta x^0 = \frac{D}{1-v} + \frac{D}{1+v} = \frac{2D}{1-v^2}.$$

According to Eq. (1.12), a clock at A' will measure a total time

$$\Delta s = (1-v^2)^{\frac{1}{2}} \Delta x^0$$

for the pulse to leave and return.

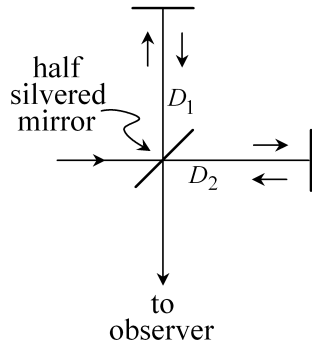
Since $c = 1$ in I' and since the light travels a distance $2D'$ in I' , Δs is also given by

$$\Delta s = 2D'.$$

Combine the last three equations to give $D = (1-v^2)^{\frac{1}{2}} D'$; the length of R' , as measured in I , is contracted by a factor $(1-v^2)^{\frac{1}{2}}$ of its rest length.

A.6 The Two Way Speed of Light. This appendix describes three experiments which together show that the two way speed of light is the same in different directions, at different times, and in different places.

The Michelson-Morley Experiment. Michelson and Morley compared the two way speed of light in perpendicular directions. They used a *Michelson interferometer* that splits a beam of monochromatic light in perpendicular directions by means of a half-silvered mirror. The beams reflect off mirrors and return to the half-silvered mirror, where they reunite and proceed to an observer. See Fig. A.10.



Let $2T$ be the time for the light to traverse an arm of the interferometer and return, D the length of the arm, and c the two way speed of light in the arm. Then by the definition of two way speed, $2T = 2D/c$. The difference in the times for the two arms is $2D_1/c_1 - 2D_2/c_2$. If f is the frequency of the light, then the light in the two arms will reunite

$$N = f \left(\frac{2D_1}{c_1} - \frac{2D_2}{c_2} \right) \quad (\text{A.1})$$

cycles out of phase.

Fig. A.10: A Michelson interferometer. If N is a whole number, the uniting beams will constructively interfere and the observer will see light; if N is a whole number $+\frac{1}{2}$, the uniting beams will destructively interfere and the observer will not see light.

Rotating the interferometer 90° switches c_1 and c_2 , giving a new phase difference

$$N' = f \left(\frac{2D_1}{c_2} - \frac{2D_2}{c_1} \right).$$

As the interferometer rotates, light and dark will alternate $N - N'$ times. Set $c_2 - c_1 = \Delta c$ and use $c_1 \approx c_2 = c$ (say) to give

$$N - N' = 2f (D_1 + D_2) \left(\frac{1}{c_1} - \frac{1}{c_2} \right) \approx 2f \frac{D_1 + D_2}{c} \frac{\Delta c}{c}. \quad (\text{A.2})$$

In Joos' experiment, $f = 6 \times 10^{14}/\text{sec}$, $D_1 = D_2 = 2100 \text{ cm}$, and $|N - N'| < 10^{-3}$. From Eq. (A.2),

$$\left| \frac{\Delta c}{c} \right| < 6 \times 10^{-12}.$$

The Brillit-Hall Experiment. In this experiment, the output of a laser was reflected between two mirrors, setting up a standing wave. The frequency of the laser was servostabilized to maintain the standing wave. This part of the experiment was placed on a granite slab that was rotated. Fig. A.11 shows a schematic diagram of the experiment.

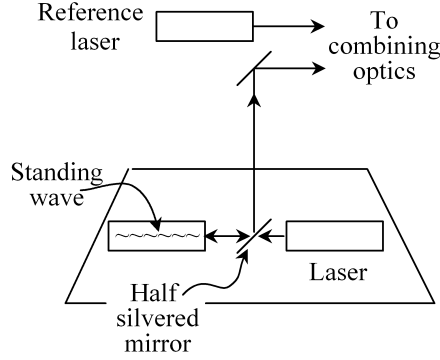


Fig. A.11: The Brillit-Hall experiment.

Thus any fractional change in c as the slab rotates would be accompanied by an equal fractional change in f . The frequency was monitored by diverting a portion of the light off the slab and comparing it with the output of a reference laser which did not rotate. The fractional change in f was no more than four parts in 10^{15} .

The Kennedy-Thorndike Experiment. Kennedy and Thorndike used a Michelson interferometer with arms of unequal length. Instead of rotating the interferometer to observe changes in N in they observed the interference of the uniting beams over the course of several months. Any change dc in the two way speed of light (due, presumably, to a change in the Earth's position or speed) would result in a change dN in N : if we set, according to the result of the Michelson-Morley experiment, $c_1 = c_2 = c$ in Eq. (A.1), we obtain

$$dN = 2f \frac{D_1 - D_2}{c} \frac{dc}{c}$$

In the experiment, $f = 6 \times 10^{14}/\text{sec}$, $D_1 - D_2 = 16 \text{ cm}$, and $|dN| < 3 \times 10^{-3}$. Thus $|dc/c| < 6 \times 10^{-9}$.

Let $2T$ be the time for light to travel from one of the mirrors to the other and back, f be the frequency of the light, and N be the number of wavelengths in the standing wave. The whole number N is held constant by the servo. Then

$$Tf = N. \quad (\text{A.3})$$

But $2T = 2D/c$, where D is the distance between the mirrors and c is the two way speed of light between the mirrors. Substitute this into Eq. (A.3) to give $Df = cN$.

A.7 Newtonian Orbits. Let the Sun be at the origin and $\mathbf{r}(t)$ be the path of the planet. At every point $\mathbf{r}$ define orthogonal unit vectors

$$\begin{aligned}\mathbf{u}_r &= \cos \theta \mathbf{i} + \sin \theta \mathbf{j}, \\ \mathbf{u}_\theta &= -\sin \theta \mathbf{i} + \cos \theta \mathbf{j}.\end{aligned}$$

See Fig. 7. Differentiate $\mathbf{r} = r\mathbf{u}_r$ to give the velocity

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{dr}{dt}\mathbf{u}_r + r\frac{d\mathbf{u}_r}{d\theta}\frac{d\theta}{dt} = \frac{dr}{dt}\mathbf{u}_r + r\frac{d\theta}{dt}\mathbf{u}_\theta.$$

A second differentiation gives the acceleration

$$\mathbf{a} = \left[\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \right] \mathbf{u}_r + \left[r^{-1} \frac{d(r^2 d\theta/dt)}{dt} \right] \mathbf{u}_\theta.$$

Since the acceleration is radial, the coefficient of $\mathbf{u}_\theta$ is zero, i.e.,

$$r^2 \frac{d\theta}{dt} = A, \quad (\text{A.4})$$

a constant. (This is Kepler's law of areas, i.e., conservation of angular momentum. Cf. Eq. (3.15).)

Using Eqs. (2.1) and (A.4), we have

$$\mathbf{a} = -\frac{\kappa M}{r^2}\mathbf{u}_r = -\frac{\kappa M}{A} \frac{d\theta}{dt}\mathbf{u}_r = \frac{\kappa M}{A} \frac{d\mathbf{u}_\theta}{dt}.$$

Integrate to obtain $\mathbf{v}$ and equate it to the expression for $\mathbf{v}$ above:

$$\frac{\kappa M}{A} (\mathbf{u}_\theta + \mathbf{e}) = \frac{dr}{dt}\mathbf{u}_r + r\frac{d\theta}{dt}\mathbf{u}_\theta,$$

where the vector $\mathbf{e}$ is a constant of integration. Take the inner product of this with $\mathbf{u}_\theta$ and use Eq. (A.4) again:

$$\frac{\kappa M}{A^2} [1 + e \cos(\theta - \theta_p)] = \frac{1}{r}, \quad (\text{A.5})$$

where $e = |\mathbf{e}|$ and $\theta - \theta_p$ is the angle between $\mathbf{u}_\theta$ and $\mathbf{e}$. This is the polar equation of an ellipse with eccentricity e and perihelion (point of closest approach to the Sun) at $\theta = \theta_p$.

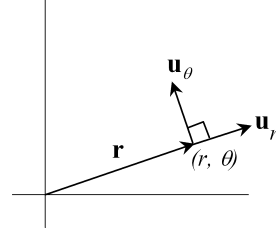


Fig. A.12: The polar unit vectors.

A.8 Geodesic Coordinates. By definition, the metric $\mathbf{f}(x) = (f_{mn}(x))$ of a local planar frame at P satisfies $(f_{mn}(P)) = \mathbf{f}^\circ$. Given a local planar frame with coordinates x , we construct a new local planar frame with coordinates $\bar{x}$ in which the metric $(\bar{f}_{mn}(\bar{x}))$ additionally satisfies $\partial_i \bar{f}_{mn}(P) = 0$. The coordinates are called *geodesic coordinates*.

We first express the derivatives of the metric in terms of the Christoffel symbols by “inverting” the definition Eq. (2.17) of the latter in terms of the former. Multiply Eq. (2.17) by (g_{ji}) and use the fact that $(g_{ji}g^{im})$ is the identity matrix to obtain $g_{ji}\Gamma_{kp}^i = \frac{1}{2}(\partial_p g_{kj} + \partial_k g_{jp} - \partial_j g_{kp})$. Exchange j and k to obtain $g_{ki}\Gamma_{jp}^i = \frac{1}{2}(\partial_p g_{jk} + \partial_j g_{kp} - \partial_k g_{jp})$. Add the two equations to obtain the desired expression:

$$\partial_p g_{jk} = g_{ji}\Gamma_{kp}^i + g_{ki}\Gamma_{jp}^i. \quad (\text{A.6})$$

Define new coordinates $\bar{x}$ by

$$x^m = \bar{x}^m - \frac{1}{2}\Gamma_{st}^m \bar{x}^s \bar{x}^t,$$

where the Christoffel symbols are those of the x -coordinates at P . Differentiate this and use $\Gamma_{it}^m = \Gamma_{ti}^m$:

$$\frac{\partial x^m}{\partial \bar{x}^j} = \frac{\partial \bar{x}^m}{\partial \bar{x}^j} - \frac{1}{2}\Gamma_{sj}^m \bar{x}^s - \frac{1}{2}\Gamma_{jt}^m \bar{x}^t = \delta_j^m - \Gamma_{sj}^m \bar{x}^s.$$

Substitute this into Eq. (2.9):

$$\bar{f}_{jk}(\bar{x}) = f_{jk}(x) - f_{jn}(x)\Gamma_{sk}^n \bar{x}^s - f_{mk}(x)\Gamma_{sj}^m \bar{x}^s + 2^{\text{nd}} \text{ order terms}.$$

This shows already that the $\bar{x}$ are inertial frame coordinates: $\bar{\mathbf{f}}(P) = \mathbf{f}(P) = \mathbf{f}_0$. Differentiate the above equation with respect to $\bar{x}^p$, evaluate at P , and apply Eq. (A.6) to $\mathbf{f}$:

$$\begin{aligned} \partial_p \bar{f}_{jk} &= \partial_s f_{jk}(P) \frac{\partial x^s}{\partial \bar{x}^p} - (f_{jn}(P)\Gamma_{pk}^n + f_{mk}(P)\Gamma_{pj}^m) \\ &= \partial_p f_{jk}(P) - (f_{jn}(P)\Gamma_{pk}^n + f_{mk}(P)\Gamma_{pj}^m) \\ &= 0. \end{aligned}$$

Thus $\bar{x}$ is a geodesic coordinate system.

A.9 The Geodesic Equations. This appendix translates the local form of the geodesic equations Eq. (2.16) to the global form Eq. (2.18). Let (f_{jk}) represent the metric with respect to a local planar frame at P with geodesic coordinates (x^i) . Then from Eq. (2.6), $(f_{jk}(P)) = \mathbf{f}^\circ$. And from Eq. (2.19), $\partial_i f_{jk}(P) = 0$.

Let (y^i) be a global coordinate system with metric $\mathbf{g}(y^i)$. We use the notation

$$y_r^i = \frac{\partial y^i}{\partial x^r}, \quad x_v^u = \frac{\partial x^u}{\partial y^v}, \quad y_{rs}^i = \frac{\partial^2 y^i}{\partial x^s \partial x^r}, \quad x_{vw}^u = \frac{\partial^2 x^u}{\partial y^w \partial y^v}.$$

Differentiate $\dot{y}^i = y_q^i \dot{x}^q$ and use Eq. (2.16):

$$\ddot{y}^i = y_q^i \ddot{x}^q + y_{qr}^i \dot{x}^r \dot{x}^q = y_{qr}^i x_j^r \dot{y}^j x_k^q \dot{y}^k.$$

In terms of components, the formula $\mathbf{g}^{-1} = \mathbf{a}^{-1} \mathbf{f}^{\circ -1} (\mathbf{a}^{-1})^t$ from Exercise 2.5c is

$$g^{im} = y_r^i f^{\circ rs} y_s^m.$$

Eq. (2.12) shows that Eq. (2.9) holds throughout the local planar frame, not just at P . Apply ∂_i to Eq. (2.9), and use Eq. (2.19) and the symmetry of $\mathbf{f}$ and $\mathbf{g}$. Evaluate at P :

$$\partial_i g_{jk} = (\partial_p f_{mn}) x_i^p x_j^m x_k^n + f_{mn} x_{ji}^m x_k^n + f_{mn} x_j^m x_{ki}^n = 2f_{mn}^{\circ} x_j^m x_{ki}^n.$$

Note that $(x_m^u y_s^m) = (\partial x^u / \partial x^s)$ is the identity matrix. Substitute the last two displayed equations in the definition Eq. (2.17) of the Christoffel symbols:

$$\begin{aligned} \Gamma_{jk}^i &= \frac{1}{2} g^{im} (\partial_k g_{jm} + \partial_j g_{mk} - \partial_m g_{jk}) \\ &= y_r^i f^{\circ rs} y_s^m (f_{uv}^{\circ} x_j^u x_{mk}^v + f_{uv}^{\circ} x_m^u x_{kj}^v - f_{uv}^{\circ} x_j^u x_{km}^v) \\ &= y_r^i f^{\circ rs} y_s^m f_{uv}^{\circ} x_m^u x_{kj}^v \\ &= y_r^i f^{\circ rs} f_{uv}^{\circ} x_m^u y_s^m x_{kj}^v \\ &= y_r^i f^{\circ rs} f_{sv}^{\circ} x_{kj}^v \\ &= y_r^i x_{kj}^r. \end{aligned}$$

Substitute for $\ddot{y}^i$ and Γ_{jk}^i from above to give the global form Eq. (2.18) of the geodesic equations:

$$\ddot{y}^i + \Gamma_{jk}^i \dot{y}^j \dot{y}^k = (y_{qr}^i x_k^q x_j^r + y_r^i x_{kj}^r) \dot{y}^j \dot{y}^k = \partial_k (y_r^i x_j^r) \dot{y}^j \dot{y}^k = 0.$$

A.10 Tensors. Tensors are a generalization of vectors. The components a_i of a vector are specified using one index. Many objects of interest have components specified by more than one index. For example, the metric g_{ij} has two. Another example is the *Riemann (curvature) tensor*, which has four:

$$R^p_{rsq} = \Gamma^p_{ts} \Gamma^t_{rq} - \Gamma^p_{tq} \Gamma^t_{rs} + \partial_s \Gamma^p_{rq} - \partial_q \Gamma^p_{rs}. \quad (\text{A.7})$$

We shall see that the metric and the Riemann tensor are tensors. A *scalar* is a tensor with a single component, i.e., it is a number, and so requires no indices.

As with a vector, we should think of a tensor as a single object, existing independently of any coordinate system, but which in a given coordinate system acquires components. There is a fully developed algebra and calculus of tensors. This formalism is a powerful tool for computations in general relativity, but we do not need it for a conceptual understanding of the theory.

A tensor index is written as a superscript (called a *contravariant* index), or as a subscript (called *covariant*). A superscript p index can be *lowered* to a subscript t index by multiplying by g_{tp} and summing. For example, $R_{trsa} = g_{tp} R^p_{rsa}$. Similarly, an index can be *raised* using g^{pt} : $R^p_{rsa} = g^{pt} R_{trsa}$. Lowering an index is a combination of two tensor operations: a *tensor product* $g_{tu} R^p_{rsa}$, followed by a *contraction* $g_{tp} R^p_{rsa}$, formed by summing over a repeated upper and lower index.

The Ricci tensor, Eq. (2.25), is a contraction of the Riemann tensor:

$$R_{jk} = R^p_{j kp}.$$

The components of a vector change under a rotation of coordinates. Similarly, the components of a tensor change under a change of coordinate systems. The formula for transforming the components from y to $\bar{y}$ coordinates can be discerned from the example of the Riemann tensor:

$$\bar{R}^a_{bcd} = R^p_{rsq} \frac{\partial \bar{y}^a}{\partial y^p} \frac{\partial y^r}{\partial \bar{y}^b} \frac{\partial y^s}{\partial \bar{y}^c} \frac{\partial y^q}{\partial \bar{y}^d}.$$

(A direct verification of this is extremely messy.) There is a derivative for each index and the derivatives are different for upper and lower indices.

We adopt the transformation law as our definition of a tensor: “a tensor is something which transforms as a tensor”. Eq. (2.12) shows that the metric is a tensor. A scalar tensor has the same value in all coordinate systems. Not all objects specified with indices form a tensor. For example, the Christoffel symbols do not.

Our rather old-fashioned definition of a tensor is not very satisfactory, as it gives no geometric insight into what a tensor is. However, it is the best we can do without getting into heavier mathematics. Exercise A.10 gives practice with the algebra of tensors.

- Exercise A.10.** a. Show that the coordinate differences (dy^i) form a tensor.
 b. Show that the product of two tensors is a tensor.
 c. Show that the contraction of a tensor is a tensor.
 d. Show that raising and lowering a tensor index are inverse operations.

Exercise A.11. This exercise shows that vectors in $\mathbb{R}^n$ are tensors. Let $(y^i) = (y^i(x))$ be curvilinear coordinates with metric g . Fix a point P .

a. Define a basis (∂_i) at P by $\partial_i = \partial \mathbf{x} / \partial y^i$. The vector ∂_i is tangent to the curve obtained by varying y^i while holding the other y 's fixed. Expand a vector $\mathbf{v}$ with respect to this basis: $\mathbf{v} = v^i \partial_i$. Show that the v^i are components of a contravariant tensor.

b. Define a second basis $(d\mathbf{y}^j)$ by $d\mathbf{y}^j = \nabla y^j$. The vector $d\mathbf{y}^j$ is orthogonal to the surface $y^j = \text{constant}$. Expand $\mathbf{v}$ with respect to this basis: $\mathbf{v} = v_j d\mathbf{y}^j$. Show that the v_j are components of a covariant tensor.

c. Show that the two bases are *dual*: $\langle \partial_i, d\mathbf{y}^j \rangle = \delta_i^j$, where $\langle \cdot, \cdot \rangle$ is the inner product.

d. Show that $v_j = g_{ij} v^i$.

e. Show that $\langle \mathbf{v}, \mathbf{w} \rangle = g_{ij} v^i w^j$.

We give only an example of the calculus of tensors. Consider the geodesic equations Eq. (2.21). The derivatives $(\dot{y}^i)$ are the components of a tensor. The second derivatives $(\ddot{y}^i)$ are not. The Christoffel symbols on the left side of the geodesic equations “correct” this: the quantities $(\ddot{y}^i + \Gamma_{jk}^i \dot{y}^j \dot{y}^k)$, are the components of a tensor, the “correct” derivative of $(\dot{y}^i)$. The right sides of the geodesic equations are zero, the components of the *zero tensor*. Thus the geodesic equations form a tensor equation, valid in all coordinate systems.

A.11 Fermi Normal Coordinates. Recall from Appendix 8 that the metric $\mathbf{f}$ of a geodesic coordinate system at E in spacetime satisfies $(f_{mn}(E)) = \mathbf{f}^\circ$ and $\partial_i f_{mn}(E) = 0$. In spacetime, a *Fermi normal coordinate system* satisfies in addition $\partial_0 \partial_i f_{mn}(E) = 0$. We do not give a construction of these coordinates.

In Fermi normal coordinates the expansion of $\mathbf{f}$ to second order at E is

$$\begin{aligned} f_{00} &= 1 - R_{0l0m} x^l x^m + \dots \\ f_{0i} &= 0 - \frac{2}{3} R_{0lim} x^l x^m + \dots \\ f_{ij} &= -\delta_{ij} - \frac{1}{3} R_{iljm} x^l x^m + \dots, \end{aligned} \quad (\text{A.8})$$

where none of i, j, l, m is 0; the R_{iljm} are components of the Riemann tensor, Eq. (A.7), at E ; and $\delta_{ij} = 1$ if $i = j$, 0 otherwise. Proof: The constant terms in the expansion express $(f_{mn}(E)) = \mathbf{f}^\circ$. The linear terms vanish because $\partial_i f_{mn}(E) = 0$. The quadratic terms do not include $l = 0$ or $m = 0$ because $\partial_0 \partial_i f_{mn}(E) = 0$. And a calculation (with a computer!) of the Riemann tensor from the metric shows that the R_{iljm} are components of that tensor. (Several symmetries of the Riemann tensor facilitate the calculation:

$$\begin{aligned} R_{iljm} &= -R_{ilmj}, \quad R_{iljm} = -R_{lijm}, \\ R_{iljm} &= R_{imjl}, \quad R_{iljm} + R_{imlj} + R_{ijml} = 0. \end{aligned}$$

The first two identities imply that $R_{ilmm} = 0$ and $R_{iijm} = 0$.)

A calculation of the Einstein tensor from the expansion Eq. (A.8) gives

$$G^{00} = R_{1212} + R_{2323} + R_{3131}. \quad (\text{A.9})$$

Eq. (2.23) gives the curvature K of a surface when $g_{12} = 0$. The general formula is

$$K = \frac{R_{1212}}{\det(\mathbf{g})}. \quad (\text{A.10})$$

Now consider the surface formed by holding the time coordinate x^0 and a spatial coordinate, say x^3 , fixed, while varying the other two spatial coordinates, x^1 and x^2 . From Eq. (A.8) the surface has metric

$$f_{ij} = \delta_{ij} + \frac{1}{3} R_{iljm} x^l x^m + \dots,$$

where i, j, l, m vary over 1, 2. A calculation shows that the 1212 component of the Riemann tensor for this metric is $-R_{1212}$. (Remember that R is the Riemann tensor for the spacetime, not the surface.) Thus from Eq. (A.10), the curvature of the surface at the origin is

$$K_{12} = \frac{-R_{1212}}{\det(\delta_{ij})} = -R_{1212}. \quad (\text{A.11})$$

Putting together Eqs. (A.9) and (A.11),

$$G^{00} = -(K_{12} + K_{23} + K_{31}).$$

A.12 The Form of the Field Equation. This Appendix (1) justifies the passage from Eq. (2.33) to Eq. (2.34) and (2) shows that $A = 1$ in Eq. (2.34).

(1) We assume that Eq. (2.33) holds at E in local inertial frames at E . From Exercise 2.16, the right side of the equation transforms contravariantly. Thus if we transform G^{jk} contravariantly to other coordinate systems, then Eq. (2.33) is valid in all coordinate systems:

$$G^{jk} = -8\pi\kappa T^{jk}. \quad (\text{A.12})$$

Is this equation Eq. (A.12) the form of the field equation in all coordinate systems? First of all, note that the expression of a physical law in one coordinate system does not uniquely determine its expression in another, just as an equation of a unit circle in cartesian coordinates (say $x^2 + y^2 = 1$) does not uniquely determine its equation in polar coordinates ($r = 1$ or $r(r+1) = r+1$ or ...). However, the validity of Eq. (A.12) in local inertial frames implies its validity in all coordinate systems. Thus there is no loss of generality in taking Eq. (A.12) as the form of the field equation in all coordinate systems.

In the text we assumed that the G^{jk} depend only on the g_{jk} and their first and second derivatives. We also assumed that $\partial T^{jk}(E)/\partial x^j = 0$ in local inertial frames at E . From Eq. (A.12), $\partial G^{jk}(E)/\partial x^j = 0$. According to a mathematical theorem of Lovelock, G^{jk} must be of the form given in the left side of Eq. (2.34).

(2) At this point in the argument on p. 48 we have already set $\Lambda = 0$. We use results from Chapter 4 with $\Lambda = 0$. Solve Eq. (4.15) for S'' and use Eq. (4.14):

$$S'' = -\frac{4\pi\kappa\rho S}{3}. \quad (\text{A.13})$$

Now consider a small ball centered at the origin. Using Eq. (4.5), the ball has radius $S(t) \int d\sigma = S(t) \sigma$ (σ is constant).

Now remove the matter from a thin spherical shell at radius $S(t) \sigma$. The shell is so thin that the removal changes the metric in the shell negligibly. An inspection of the derivation of the Schwarzschild metric in Section 3.2 shows that metric in the shell is entirely determined by the total mass inside the shell. In particular, we may remove the mass outside the shell without changing the metric in the shell. Do so.

Apply the Newtonian equation Eq. (2.1) to an inertial particle in the shell:

$$S''\sigma = -\kappa \frac{\frac{4}{3}\pi(S\sigma)^3\rho}{(S\sigma)^2}.$$

This is the same as Eq. (A.13); Einstein's and Newton's theories give the same result.

But if $A \neq 1$ in Eq. (2.34), then the left side of Eq. (A.13) would be multiplied by A . Thus the coincidence of the two theories in this situation requires $A = 1$.

A.13 Eliminating the dt dr Term. We make a coordinate change in Eq. (3.2) to eliminate the $dt dr$ term. From the theory of differential equations there is an integrating factor $I(t, r)$ and then a function $\tau(t, r)$ so that

$$d\tau = I(e^{2\mu} dt - u dr).$$

Then

$$dt = e^{-2\mu}(I^{-1}d\tau + u dr).$$

Substitute for dt and then for $d\tau$ in the $d\tau dr$ term:

$$e^{2\mu} dt^2 - 2u dt dr - e^{2\nu} dr^2 = I^{-2} e^{-2\mu} d\tau^2 - (2u^2 e^{-2\mu} + e^{2\nu}) dr^2 \equiv e^{2\bar{\mu}} d\tau^2 - e^{2\bar{\nu}} dr^2.$$

Drop the bars and rename τ to t and this becomes

$$e^{2\mu} dt^2 - e^{2\nu} dr^2.$$

There is no $dt dr$ term.

A.14 Approximating an Integral. Retaining only first order terms in the small quantities m/r and m/r_p we have

$$\begin{aligned}
\frac{(1 - 2m/r)^{-2}}{1 - \frac{1 - 2m/r}{1 - 2m/r_p} (r_p/r)^2} &\approx \frac{(1 + 2m/r)^2}{1 - (1 - 2m/r)(1 + 2m/r_p)(r_p/r)^2} \\
&\approx \frac{1 + 4m/r}{1 - (1 - 2m/r + 2m/r_p)(r_p/r)^2} \\
&= \frac{1 + 4m/r}{\left(1 - (r_p/r)^2\right) \left(1 - \frac{2mr_p}{r(r + r_p)}\right)} \\
&\approx \left(1 - (r_p/r)^2\right)^{-1} (1 + 4m/r) \left(1 + \frac{2mr_p}{r(r + r_p)}\right) \\
&\approx \left(1 - (r_p/r)^2\right)^{-1} \left(1 + 4m/r + \frac{2mr_p}{r(r + r_p)}\right).
\end{aligned}$$

We now see that Eq. (3.25) can be replaced with

$$t = \int_{r_e}^{r_p} \left(1 - \left(\frac{r_p}{r}\right)^2\right)^{-1} \left(1 + \frac{4m}{r} + \frac{2mr_p}{r(r + r_p)}\right) dr.$$

Integrate to give

$$t = (r_e^2 - r_p^2)^{\frac{1}{2}} + 2m \ln \frac{r_e + (r_e^2 - r_p^2)^{\frac{1}{2}}}{r_p} + m \left(\frac{r_e - r_p}{r_e + r_p}\right)^{\frac{1}{2}}.$$

Use $r_p \ll r_e$ to obtain Eq. (3.26).

A.15 The Robertson-Walker Metric. This appendix derives the Robertson-Walker for an isotropic universe, Eq. (4.4).

In Chapter 3 we showed that the metric of a spherically symmetric spacetime can be put in the form Eq. (3.2):

$$ds^2 = e^{2\mu} dt^2 - e^{2\nu} dr^2 - r^2 e^{2\lambda} d\Omega^2, \quad (\text{A.14})$$

where the coefficients do not depend on ϕ or θ .

Since the coordinates are comoving, $dr = d\phi = d\theta = 0$ for neighboring events on the worldline of a galaxy. Thus by Eq. (A.14), $ds^2 = e^{2\mu} dt^2$. But $ds = dt$, since both are the time between the events measured by a clock in the galaxy. Thus $e^{2\mu} = 1$. Our metric is now of the form

$$ds^2 = dt^2 - (e^{2\nu} dr^2 - r^2 e^{2\lambda} d\Omega^2) \quad (\text{A.15})$$

If the universe is expanding isotropically about every galaxy, then it seems that the spatial part of the metric could depend on t only through a common factor, say $S(t)$, as in the metric of our balloon, Eq. (4.2). We now demonstrate this.

Consider light sent from a galaxy at (t, r, ϕ, θ) to a galaxy at $(t + dt, r + dr, \phi, \theta)$. From Eq. (A.15), $dt^2 = e^{2\nu} dr^2$. Thus the light travel time between $P(r, \phi, \theta)$ and $Q(r + dr, \phi, \theta)$ is $e^{\nu(t,r)} dr$. The ratio of this time to that measured at another time t' is given by $e^{\nu(t,r)}/e^{\nu(t',r)}$. Similarly, the ratio of the times between P and $R(r, \theta, \phi + d\phi)$ is $e^{\lambda(t,r)}/e^{\lambda(t',r)}$. By isotropy at P the ratios are equal:

$$\frac{e^{\nu(t,r)}}{e^{\nu(t',r)}} = \frac{e^{\lambda(t,r)}}{e^{\lambda(t',r)}}. \quad (\text{A.16})$$

(For example, if the time from the galaxy at P to the nearby galaxy at Q doubles from t to t' , then the time from P in another direction to the nearby galaxy at R must also double.)

The two sides of Eq. (A.16) are independent of r . For if two r 's gave different ratios, then there would not be isotropy half way between. Fix t' . Then both members of Eq. (A.16) are a function only of t , say $S(t)$. Set $\nu(t', r) = \nu(r)$ to obtain $e^{\nu(t,r)} = S(t) e^{\nu(r)}$. Similarly, $e^{\lambda(t,r)} = S(t) e^{\lambda(r)}$. Thus the metric Eq. (A.15) becomes

$$ds^2 = dt^2 - S^2(t) (e^{2\nu} dr^2 + r^2 e^{2\lambda} d\Omega^2). \quad (\text{A.17})$$

The coordinate change $\bar{r} = r e^{\lambda(r)}$ puts Eq. (A.17) in the form

$$ds^2 = dt^2 - S^2(t) (e^{2\nu} dr^2 + r^2 d\Omega^2) = dt^2 - S^2(t) d\sigma^2. \quad (\text{A.18})$$

We now determine ν in Eq. (A.18).

Exercise A.12. Show that an application of Eq. (2.23) shows that the curvature K of the half plane $\theta = \theta_0$ in the $d\sigma$ portion of the metric satisfies $\nu' e^{-2\nu} = Kr$.

As above, K is independent of r . Integrating thus gives $e^{-2\nu} = C - Kr^2$, where C is a constant of integration.

It only remains to determine C . From Eq. (A.18) we can label the sides of an infinitesimal sector of an infinitesimal circle centered at the origin in the $\phi = \pi/2$ plane. See Fig. A.13. Since the length of the subtended arc is the radius times the subtended angle, $e^{\nu(0)} = 1$. Thus $C = 1$.

Set $k = 1, 0, -1$ according as $K > 0, K = 0, K < 0$. Then k indicates the sign of the curvature of space. If $K \neq 0$, substitute $\bar{r} = r |K|^{\frac{1}{2}}$ and $\bar{S} = S |K|^{\frac{1}{2}}$. Dropping the bars, we obtain

$$ds^2 = dt^2 - S^2(t) \left(\frac{dr^2}{1 - k r^2} + r^2 d\Omega^2 \right) \quad (k = 0, \pm 1).$$

The substitution $r = \bar{r}/(1 + k\bar{r}^2/4)$ gives the Robertson-Walker metric.

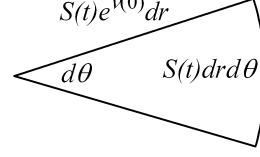


Fig. A.13: Evaluating the integration constant C .

A.16 Redshift Relations. This appendix relates three physical parameters of a distant object to its redshift. We obtain these relationships for a flat universe ($k = 0$) with a cosmological constant Λ , as described in Sec. 4.4.

The Distance-Redshift Relation. The distance-redshift relation $d(z)$ expresses the distance d to a distant object today as a function of its redshift.

To obtain the relation, suppose that light is emitted toward us at event (t_e, r_e) and received by us today at event $(t_o, 0)$. Evaluate Eq. (4.14) at t_o , use Eq. (4.9), and rearrange: $8\pi\kappa\rho(t_o) = 3H_o^2 - \Lambda$. Thus from Eqs. (4.18) and (4.11),

$$8\pi\kappa\rho(t) = 8\pi\kappa\rho(t_o) \frac{S^3(t_o)}{S^3(t)} = (3H_o^2 - \Lambda)(z+1)^3.$$

Thus Eq. (4.14) gives

$$\left(\frac{S'}{S}\right)^2 = \frac{8\pi\kappa\rho(t)}{3} + \frac{\Lambda}{3} = H_o^2 [\Omega_\Lambda + (1 - \Omega_\Lambda)(z+1)^3], \quad (\text{A.19})$$

where $\Omega_\Lambda \equiv \Lambda/3H_o^2$. (Using Eq. (4.17), Eq. (4.16) can be written $\rho/\rho_c + \Omega_\Lambda = 1$. Thus Ω_Λ is the dark energy fraction of the matter-energy density of the universe.)

For light emitted at time t and received by us at t_o , Eq. (4.11) can be written $S(t_o) = (z(t) + 1)S(t)$. Differentiate and substitute into Eq. (A.19):

$$(1+z)^{-2} \left(\frac{dz}{dt}\right)^2 = H_o^2 [\Omega_\Lambda + (1 - \Omega_\Lambda)(z+1)^3]. \quad (\text{A.20})$$

From Eq. (4.10) with $k = 0$, $dt = S(t)dr = S(t_o)(1+z)^{-1}dr$. And from Eqs. (4.8) and (4.4) with $ds = 0$ and $k = 0$, $d = S(t_o)r_e$. Substitute these into Eq. (A.20), separate variables, and integrate to give the distance-redshift relation:

$$d(z) = S(t_o)r_e = H_o^{-1} \int_0^z [\Omega_\Lambda + (1 - \Omega_\Lambda)(\zeta+1)^3]^{-\frac{1}{2}} d\zeta. \quad (\text{A.21})$$

The Angular Size-Redshift Relation. The angular size-redshift relation $\alpha(z)$ expresses the angular size α of a distant object as a function of its redshift.

To obtain the relation, send a pulse of light from one side of the object to the other. Let $(t_e, r_e, \theta_e, \phi_e)$ and $(t_e + dt, r_e, \theta_e, \phi_e + d\phi)$ be the coordinates of the emission and reception events. From Eq. (4.4), $dt = r_e S(t_e) d\phi$. Since $c = 1$, the size of the object $D = dt$. Also, $\alpha = d\phi$. Thus using Eq. (4.11),

$$\alpha = \frac{D}{r_e S(t_e)} = \frac{(z+1)D}{r_e S(t_o)}.$$

Substitute from the distance-redshift relation, Eq. (A.21), to obtain the angular size-redshift relation:

$$\alpha(z) = \frac{(z+1)H_o D}{\int_0^z [\Omega_\Lambda + (1 - \Omega_\Lambda)(\zeta + 1)^3]^{-\frac{1}{2}} d\zeta}. \quad (\text{A.22})$$

The Luminosity-Redshift Relation. The luminosity-redshift relation $\ell(z)$ expresses the luminosity ℓ of light received from a distant object as a function of its redshift.

To obtain the relation, suppose that light of intrinsic luminosity L is emitted toward us at event (t_e, r_e) and received by us at event $(t_o, 0)$. Replace the approximate relation $\ell = L/4\pi d^2$ from Sec. 4.1 with this exact relation:

$$\ell = \frac{L}{4\pi(1+z)^2 d^2}. \quad (\text{A.23})$$

The new $1+z$ factors are most easily understood using the photon description of light. The energy of a photon is $E = hf$ (an equation discovered by Einstein), where h is *Planck's constant*. According to Eq. (1.7), the energy of each received photon is diminished by a factor $1+z$. And according to Eq. (1.6), the rate at which photons are received is diminished by the same factor.

Substitute d from the distance-redshift relation, Eq. (A.21), into Eq. (A.23) to obtain the luminosity-redshift relation:

$$\ell(z) = \frac{H_o^2 L}{4\pi(1+z)^2 \left(\int_0^z \{\Omega_\Lambda + (1 - \Omega_\Lambda)(\zeta + 1)^3\}^{-\frac{1}{2}} d\zeta \right)^2}. \quad (\text{A.24})$$

Index

- accelerometer, [13](#), [30](#)
- accretion disk, [61](#)
- angular size, [73](#)
- angular size-redshift relation, [73](#), [100](#)
- approximating an integral, [96](#)
- approximations, [81](#)
- balloon analogy, [66](#), [67](#), [69](#), [70](#)
- big bang nucleosynthesis, [68](#), [75](#), [80](#)
- black hole, [50](#), [63](#)
- blackbody radiation, [68](#), [72](#)
- Boltzmann's constant, [72](#)
- Braginsky, V. B., [31](#)
- Brecher, K., [24](#)
- Bridgeman, P.W., [17](#)
- Brillit-Hall experiment, [23](#), [87](#)
- Christoffel symbols, [40](#)
- clock, [11](#)
- comoving coordinates, [69](#)
- contravariant index, [47](#), [91](#)
- coordinate singularity, [63](#)
- coordinates, [14](#), [34](#)
- Copernicus, Nicholas, [28](#)
- cosmic background radiation (CBR), [68](#), [72–74](#)
- cosmological constant, [74](#)
- cosmology, [66](#)
- covariant index, [46](#), [91](#)
- crab nebula, [50](#)
- critical density, [74](#)
- curvature, [43](#), [47](#), [93](#)
- curvature scalar, [45](#)
- curve, [12](#)
- curved spacetime, [11](#)
- dark energy, [75](#), [77](#), [80](#)
- dark matter, [75](#), [77](#), [80](#)
- degenerate electron pressure, [49](#)
- deSitter, Willem, [59](#)
- deuterium, [75](#)
- Dicke, R., [31](#)
- dipole anisotropy, [68](#)
- distance-redshift relation, [67](#), [71](#), [78](#), [99](#), [100](#)
- double quasar, [58](#)
- drag free satellite, [61](#)
- dust, [45](#), [74](#)
- Eddington, Sir Arthur Stanley, [57](#)
- Einstein summation convention, [37](#)
- Einstein tensor, [45](#), [93](#)
- Einstein, Albert, [11](#), [29](#), [30](#)
- eliminating the dt dr term, [95](#)
- energy-momentum tensor, [45](#), [74](#)
- ergosphere, [64](#)
- event, [11](#)
- event horizon, [78](#)
- Fermi normal coordinates, [47](#), [93](#)
- field equation, [43](#), [44](#), [46](#), [51](#), [73](#)
- flat spacetime, [11](#)
- form of the field equation, [94](#)
- frame dragging, [60](#)
- galaxy, [66](#)
- galaxy cluster, [66](#)
- Galilei, Galileo, [28](#)
- Galle, J., [28](#)
- Gauss, C. F., [30](#), [44](#)
- general relativity, [11](#)
- geodesic, [25](#), [26](#), [40–42](#)
- geodesic coordinates, [41](#), [47](#), [89](#), [90](#)
- geodesic equations, [53](#), [90](#)

geodesic postulate, 13, 25, 26, 30, 31, 40–42
 geodetic precession, 59, 61
 global coordinate postulate, 35, 36
 global coordinates, 34, 36
 GPS, 33
 gravitational binding energy, 31
 gravitational lens, 58, 72, 74, 75
 gravitational lens statistics, 74, 76
 gravitational waves, 62, 80
 gravitomagnetic clock effect, 61
 gravitomagnetism, 60
 Gravity Probe B, 60

 Hafele-Keating experiment, 12, 22, 23, 32, 33, 52
 Hils-Hall experiment, 23
 horizon, 63, 64
 Hubble’s constant, 66, 71
 Hubble’s law, 66, 67, 71
 Hubble, Edwin, 67

 inertial force, 13, 30
 inertial frame, 13, 14
 inertial frame postulate, 13, 14, 17, 30, 36
 inertial object, 13, 30, 66
 inflation, 73, 80
 integrated Sachs-Wolfe effect, 76
 intrinsic description, 36, 46
 intrinsic luminosity, 67
 isotropic universe, 69

 Joos, G., 23, 86

 Kennedy-Thorndike experiment, 23, 87
 Kepler, Johannes, 28
 Kerr metric, 60, 64

 LAGEOS satellites, 60
 Le Verrier, U., 28
 length contraction, 20, 32, 48, 84
 Lense-Thirring effect, 60
 light deflection, 57
 light delay, 58
 light retardation, 62

 light second, 19
 light speed, 24, 39
 lightlike separated, 18, 19
 local inertial frame, 35
 local planar frame, 34, 41
 loop quantum gravity, 80
 Lovelock’s theorem, 48, 94
 luminosity, 67
 luminosity-redshift relation, 75, 100
 lunar laser experiment, 31, 59

 Macek-Davis experiment, 17, 82
 manifold, 35
 mass, 48
 Mercury, 57
 metric postulate, 13, 18, 19, 22, 30, 37–39
 Michelson interferometer, 86, 87
 Michelson-Morley experiment, 23, 86
 Milky Way, 66
 Minkowski, Hermann, 19
 muon, 23

 neutron star, 50, 53, 62, 65
 Newton, Isaac, 28
 newtonian gravitational constant, 28, 52
 newtonian orbits, 88

 Painlevé coordinates, 63
 particle horizon, 78
 periastron advance, 62
 perihelion advance, 56
 photon, 100
 physical constants, 81
 planar frame, 14
 planar frame postulate, 14, 34
 Planck’s constant, 72, 100
 Planck’s law, 72
 point, 11
 postulates, global form, 36
 postulates, local form, 36
 Pound, R. V., 32
 principle of relativity, 84
 proper distance, 18, 83

proper time, 18, 20, 39
 pseudosphere, 43, 44
 Ptolemy, Claudius, 27
 pulsar, 50, 62

 quantum theory, 80
 quasar, 50, 57–59, 65, 72, 75

 Rebka, G. A., 32
 redshift, 16
 Doppler, 17, 20, 24, 32
 expansion, 17, 67, 69, 71
 gravitational, 17, 32, 33, 52, 62, 64
 Ricci tensor, 45, 91
 Riemann tensor, 91, 93
 Riemann, G. B., 36
 Robertson-Walker metric, 70, 73, 97
 rod, 11, 13, 14, 21, 34, 35, 48, 84
 round trip light speed, 82

 scalar, 91
 Schwarzschild metric, 52
 Schwarzschild radius, 63
 Schwarzschild, Karl, 52
 Sirius, 49, 53
 Sloan Digital Sky Survey (SDSS), 74
 Snider, J.L., 32
 spacelike separated, 18, 20, 83
 spacetime, 11, 19
 spacetime diagram, 15
 special relativity, 11
 standard model, 80
 static limit, 64
 stellar evolution, 49
 STEP, 31
 string theory, 80
 supernova, 49, 50, 62, 65, 68, 72, 75
 surface, 11
 surface dwellers, 13
 synchronization, 14, 15, 17

 tensors, 46, 91
 terrestrial redshift experiment, 32, 35, 53

 tidal acceleration, 35, 44
 time dilation, 20, 23, 32, 52, 62
 timelike separated, 18, 20, 21

 universal light speed, 19, 21, 23, 39
 universal time, 12, 29

 vacuum field equation, 46, 51

 white dwarf, 49, 53, 65
 WMAP, 74
 worldline, 12