General Relativity in a Nutshell

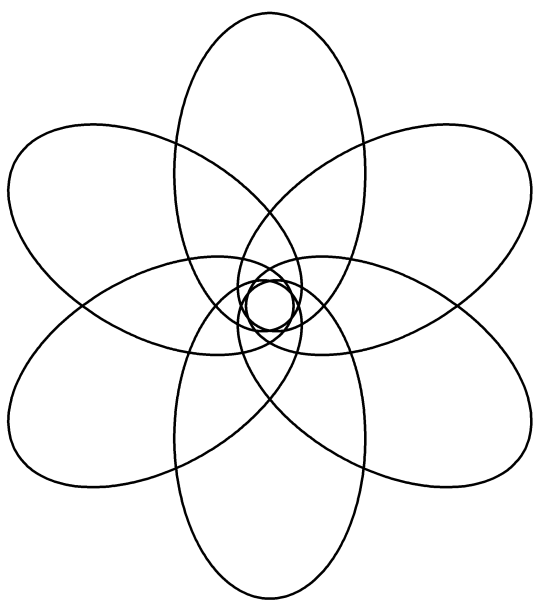
October 2, 2017

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

The purpose of this little book is to provide a clear and careful account of general relativity with a minimum of mathematics. The book has fewer prerequisites than other texts, and less mathematics is developed. The prerequisites are single variable calculus, a few basic facts about partial derivatives and line integrals, and a little matrix algebra. A little knowledge of physics is useful but not essential. The algebra of tensors plays only a minor role.[1] (Similarly, many elementary differential geometry texts develop the intrinsic geometry of curved surfaces without using tensors.)

Despite the book's brevity and modest prerequisites, it is a serious introduction to the theory and applications of general relativity which demands careful study. It can be used as a textbook for general relativity or as an adjunct to standard texts. It is also suitable for self-study by more advanced students.

All this should make general relativity available to a wider audience than before. A well prepared sophomore can learn about exciting current topics such as curved spacetime, black holes, the big bang, dark energy, and the accelerating universe.

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. This approach to the metric provides easier access to and deeper understanding of special and general relativity, and facilitates the transition from special to general relativity.

[1] The notion of covariant and contravariant indices is developed and used in one page of the text.

Chapter 2 introduces the four fundamental principles of general relativity as postulates. The purpose of the postulates is not to achieve rigor – which is neither desirable nor possible in a book at this level – but to state clearly the principles, and to exhibit clearly the relationship to special relativity and the analogy with surfaces. The first three principles are expressed in terms of local inertial frames, which tells us their physical meaning. They are then translated to global coordinates, which is necessary for calculations. The fourth principle, the field equation, is then motivated and stated. Finally, an interesting consequence of the field equation relating curvatures and density is obtained.

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, which includes gravitomagnetism and the Gravity Probe B experiment. The chapter closes with sections on the double 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 the field equation. We review the evidence for a spatially flat universe with a cosmological constant. We then apply the field equation with a cosmological constant to a spatially flat Robertson-Walker spacetime. The solution is given in closed form. Recent astronomical data allow us to specify all parameters in the solution, giving the new "standard model" of the universe with dark matter, dark energy, and an accelerating expansion.

Many recent spectacular astronomical discoveries and observations are relevant to general relativity. They are described at appropriate places in the book.

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. 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. Albert Einstein created the theory during the decade following the publication of his special theory of relativity in 1905. The special theory is a theory of space and time which does not take gravity into account. The general theory, published in 1915, generalizes the special theory to include gravity. It is commonly felt to be the most beautiful of all physical theories.

We will explore many fascinating aspects of relativity, such as the behavior of moving clocks, curved spacetimes, Einstein's field equation, the perihelion advance of Mercury, black holes, the big bang, the accelerating expansion of the universe, and dark energy. Let us begin!

Fig. 1.1: Albert Einstein, 1879-1955.

Definitions. We give definitions of three fundamental concepts used throughout this book: event, spacetime, and worldline.

Event. 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. It has neither temporal nor spatial extension. For example, the collision of two particles is an event. To be present at the event, you must be at the right place at the right time.

Spacetime. 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. Chapter 3 examines the spacetime around the Sun. We will study the motion of planets and light in the spacetime. Chapter 4 examines a spacetime for the entire universe! We will study the origin and evolution of the universe.

A *flat spacetime* is a spacetime without significant gravity. Special relativity describes flat spacetimes. A *curved spacetime* is a spacetime with significant gravity. General relativity describes curved spacetimes.

There is nothing mysterious about the words "flat" or "curved" attached to a spacetime. 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.* The analogy will be a major theme of this book: we will use our intuitive understanding of flat and curved surfaces to guide our understanding of flat and curved spacetimes.

Worldline. A curve in a surface is a continuous succession of points in the surface. A *worldline* in a spacetime is a continuous succession of events in the 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. A planet has a worldline in the curved spacetime around the Sun. Note that if the planet returns to a specific point in space during its orbit, then it does not return to the same event in the spacetime, as it returns at a later time.

Time. Relativity theory contradicts everyday views of time. The most direct illustration of this is the *Hafele-Keating experiment*, which we now describe.

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! In 1971 J. C. Hafele and R. Keating brought two atomic clocks together, placed one of them in an airplane which circled the Earth, and then brought the clocks together again. Thus the clocks had different worldlines between the event of their separation and the event of their reunion. *The clocks 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.

We shall see that relativity predicts the measured difference. We shall also see that relativity predicts large differences between clocks whose relative speed is close to that 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 were 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. But clocks traveling together between the events will measure the same time between the events. (Unless some adverse physical condition affects a clock's rate.) Thus we can speak of the time *along a given worldline* between the events.

The word "clock" here must be understood in the general sense of any physical process, e.g., the vibrations of a tuning fork, radioactive decays, or the aging of an organism. Twins, each remaining close to one of the clocks during the Hafele-Keating experiment, would age according to their clock. They would thus be of slightly different ages when reunited.

The next three sections put forward 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 (which we will define) and light move in a straight line at constant speed in inertial frames.

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

Planar Frames. 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. Surface dwellers call the coordinate system a *planar frame*. Fig. 1.3 shows the axes of two different planar frames.

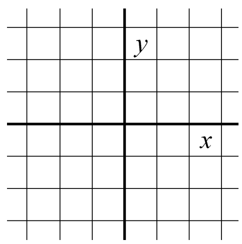


Fig. 1.2: A planar frame.

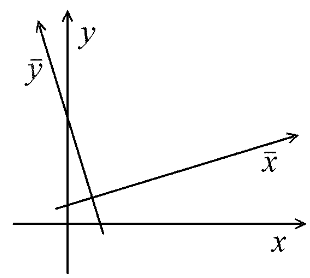


Fig. 1.3: Two planar frames.

The Planar Frame Postulate for a Flat Surface

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

Inertial Frames. Imagine that you are an astronaut in interstellar space, where gravity is insignificant. If your rocket is not firing and your ship is not spinning, then you will feel no forces acting on you and you can float freely in your cabin. If your spaceship is accelerating, then you will feel a force pushing you back against your seat. If the ship turns to the left, then you will feel a force to the right. If the ship is spinning, you will feel a force outward from the axis of spin. Call these forces *inertial forces*.

Accelerometers measure inertial forces. Fig. 1.4 shows a simple accelerometer consisting of a weight held at the center of a cubical frame by identical springs. Inertial forces cause the weight to move from the center.

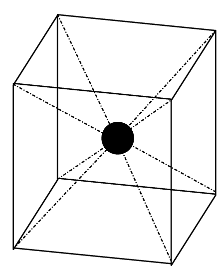


Fig. 1.4: A simple accelerometer. Acceleration causes the weight to move from the center.

An object which experiences no inertial forces is called an *inertial object*.

It is useful to label the events in a flat spacetime with coordinates (t, x, y, z). The coordinates specify *when* (the t coordinate) and *where* (the (x, y, z) coordinates) the event occurs, i.e., they completely specify the location of the event in the spacetime. 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. 1.5. Assign rectangular coordinates (x, y, z) to the nodes of the lattice in the usual way. 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.

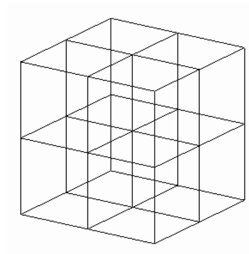


Fig. 1.5: An inertial lattice.

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

If you are an inertial object then you can construct an inertial frame in which you are at rest. If fact, you can do this in many ways:

The Inertial Frame Postulate for a Flat Spacetime

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

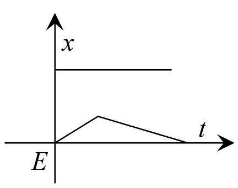


Fig. 1.6: Worldlines.

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. 1.6 shows the worldlines of two particles. One is at rest on the x-axis and the other moves away from $x = 0$ and then returns more slowly.

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.[1]

Exercise 1.2. Describe the worldline of an object moving in a circle in the $z = 0$ plane at constant speed. You need three dimensions for this, two of space and one of time.

[1] Spacetime diagrams are usually constructed with t as the vertical axis and x the horizontal axis. Then the slope is $1/v$.

Synchronization. We return to the matter of synchronizing the clocks of an inertial frame. We can directly compare side-by-side clocks to see if they are synchronized. But 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

$$t_Q = t_P + D/v. \tag{1.1}$$

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. For 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

$$t_Q - t_P = t'_P - t'_Q, \tag{1.2}$$

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 into Eq. (1.2):

$$t_Q = t_P + T; \tag{1.3}$$

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
>
> $$t'_S - t'_R = t_S - t_R. \tag{1.4}$$

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

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.

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". Appendix 3 proves this with the aid of an auxiliary assumption.

Redshifts. A rearrangement of Eq. (1.4) gives

$$\Delta s_o = \Delta s_e, \tag{1.5}$$

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.

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

$$\boxed{z = \frac{\Delta s_o}{\Delta s_e} - 1.} \tag{1.6}$$

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.

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.

Redshifts are often expressed in terms of frequencies. 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 the equation can be written

$$z = \frac{f_e}{f_o} - 1. \tag{1.7}$$

1.3 The Metric Postulate

Flat Surface Metric. Let P and Q be points in a flat surface. Different curves between the points have different lengths. But surface dwellers single out the straight line distance Δs and call it the *proper distance* between the points.

This definition of proper distance is in *geometric* terms, without using coordinates. If a planar frame is introduced, then the Pythagorean theorem gives a simple formula for Δs:

The Metric Postulate for a Flat Surface

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

$$\Delta s^2 = \Delta x^2 + \Delta y^2. \tag{1.8}$$

Here Δs^2 means $(\Delta s)^2$, Δx^2 means $(\Delta x)^2$, etc.

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. (1.8) will always produce Δs. Neither Δx nor Δy alone has a geometric significance. Together they do: they determine Δs, which has a direct geometric significance, independent of any coordinate system.

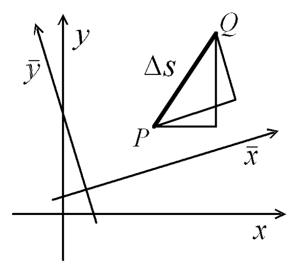


Fig. 1.7: $\Delta s^2 = \Delta x^2 + \Delta y^2 = \Delta \bar{x}^2 + \Delta \bar{y}^2$.

Flat Spacetime Metric. Let E and F be events in a flat spacetime. There is a distance-like quantity Δs between E and F, called the *interval* between them. The definition of Δs in a flat spacetime is more complicated than in a flat surface, because 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 $\Delta s = 0$ for E and F. (The reason for this peculiar definition will become clear later.)

- 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 inertial clock. This is the *proper time* between E and F. (Clocks on other worldlines between the events will measure different times.)

- If E and F can be simultaneously at the ends of an inertial rod – simultaneously in the sense that light flashes emitted at E and F reach the center of the rod simultaneously – they are *spacelike separated*. Then define $|\Delta s|$ to be the length the rod. (The reason for the absolute value will become clear in Eq. (1.12).) This is the *proper distance* between E and F.

This definition of the spacetime interval Δs just given is in *physical* terms, without using coordinates. If an inertial frame is introduced, then the metric postulate for a flat spacetime asserts a simple formula for the interval:

The Metric Postulate for a Flat Spacetime

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

$$\boxed{\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2.} \tag{1.9}$$

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 $\Delta x = \Delta y = \Delta z = 0$. In an inertial frame in which the clock is moving, at least one of Δx, Δy, Δz will not be zero. However, the particular combination of the differences in Eq. (1.9) will always produce Δs. None of the coordinate differences has a physical significance independent of the particular inertial frame chosen. Together they 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." The new relationship between space and time in special relativity is a radical change from the earlier view that space and time are independent.

Minkowskian geometry, with mixed signs in its expression for the interval, Eq. (1.9), is different from Euclidean geometry. But it is a perfectly valid geometry, the geometry of a flat spacetime. It will, however, take you a while to get used to it.

Physical Meaning of the Metric Postulate. We discuss the physical meaning of the metric postulate formula, Eq. (1.9). We do not need the y- and z-coordinates for this, and so we use the formula in the form

$$\Delta s^2 = \Delta t^2 - \Delta x^2. \tag{1.10}$$

We treat lightlike, timelike, and spacelike separated events individually.

Lightlike Separated. By definition, a pulse of light can move between lightlike separated events, and $\Delta s = 0$ for the events. From Eq. (1.10) the speed of the pulse is $|\Delta x|/\Delta t = 1$. *The metric postulate asserts that the speed c of light always has the value $c = 1$ in all inertial frames.*

The assertion is that the speed is the same in all inertial frames; the actual value $c = 1$ is 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 – a light year.) Then 1 cm $= 3.3 \times 10^{-11}$ sec. With this convention $c = 1$, and all other speeds are expressed as a dimensionless fraction of the speed of light. Ordinarily the fractions are *very* small. For example, 3 km/sec = .00001.

Suppose Ellen is walking in a moving train. Her speed in a rest frame of the train (i.e., an inertial frame in which the train is at rest) is different from her speed in a rest frame of the ground. But according to the metric postulate, a pulse of light is different: it has the same speed in the two frames. You must understand this difference to feel the full force of the postulate.

Timelike Separated. By definition, an inertial clock can move between timelike separated events, and it measures the proper time Δs between the events. The speed of the clock is $v = |\Delta x|/\Delta t$. From Eq. (1.10),

$$\Delta s = (\Delta t^2 - \Delta x^2)^{\frac{1}{2}} = [1 - (\Delta x/\Delta t)^2]^{\frac{1}{2}}\Delta t = (1 - v^2)^{\frac{1}{2}}\Delta t. \tag{1.11}$$

Thus $\Delta s \le \Delta t$. This is called *time dilation*. Informally, "moving clocks run slowly". Fig. 1.8 shows the graph of the time dilation factor $\Delta t/\Delta s = (1 - v^2)^{-\frac{1}{2}}$. For normal speeds, v is very small, v^2 is even smaller, and so $\Delta s \approx \Delta t$, as expected. But as $v \to 1$, $(1 - v^2)^{-\frac{1}{2}} \to \infty$.

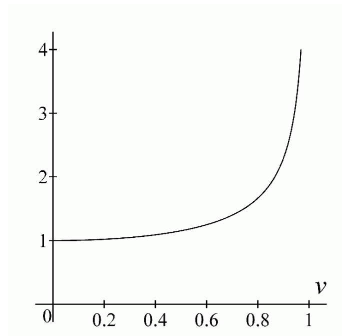


Fig. 1.8: The time dilation factor $\Delta t/\Delta s = (1 - v^2)^{-\frac{1}{2}}$ vs. v.

Exercise 1.6. Investigate the Doppler redshift. A source of light pulses moves with speed 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 their reception by the observer.

a. Show that $\Delta t_o = \Delta t_e + v\Delta t_e$.

b. Ignore time dilation in Eq. (1.6) by setting $\Delta s = \Delta t$. Show that $z = v$ in this approximation.

c. Show that the exact formula is $z = [(1+v)/(1-v)]^{\frac{1}{2}} - 1$. Use the result of part a.

Spacelike Separated. By definition, the ends of an inertial rigid rod of proper length $|\Delta s|$ can be simultaneously present at spacelike separated events. Appendix 5 shows that $\Delta t < |\Delta x|$ and that the speed of the rod in I is $v = \Delta t/|\Delta x|$ (that is not a typo).

Exercise 1.7. Show that a calculation similar to Eq. (1.11) gives

$$|\Delta s| = (1 - v^2)^{\frac{1}{2}}|\Delta x|. \tag{1.12}$$

(Since $\Delta s^2 < 0$ we must take the absolute value $|\Delta s|$.) Thus $|\Delta s| \le |\Delta x|$.

Appendix 6 discusses a different phenomenon, called *length contraction*, of moving rods.

Connections. We have just seen that the physical meaning of the metric postulate is different for lightlike, timelike, and spacelike separated events. Strange as the individual meanings might seem, they are logically connected: the meaning for lightlike separated events ($c = 1$) implies those 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 5.

Let timelike separated events E and F have coordinate differences $(\Delta t,\ \Delta x)$ in an inertial frame I. An inertial clock C moving between the events measures the proper time Δs between them. We show that if $c=1$ in all inertial frames, then $\Delta s^2 = \Delta t^2 - \Delta x^2$. Time dilation, Eq. (1.11), follows.

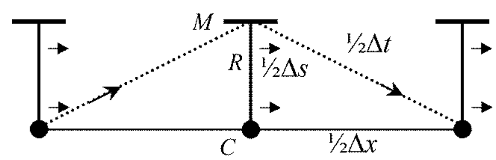


Fig. 1.9: $\Delta s^2 = \Delta t^2 - \Delta x^2$ for timelike separated events. The path of the light pulse in I is dotted.

C is attached to an end of a rod R oriented perpendicular to its motion in I. At E send a light pulse from C along R toward a mirror M attached to the other end of R. Choose the length of R so that the pulse is reflected back to C at F. From C's point of view, R is at rest and the light simply travels up it and returns. In I, the light traverses the dotted path shown in Fig. 1.9. The figure also shows the clock, rod, and mirror at three different times: when the light is sent from C, reflects off M, and returns to C.

Refer to the rightmost triangle in Fig. 1.9. Since $c = 1$ in I, the light travels a distance Δt in I. This gives the labeling $\frac{1}{2}\Delta t$ 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 $\frac{1}{2}\Delta s$ 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 6 discusses this.) Apply the Pythagorean theorem to the triangle to obtain $\Delta s^2 = \Delta t^2 - \Delta x^2$. This completes the proof.

In short, since the light travels farther in I than in I' (twice the hypotenuse vs. twice the altitude) 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.

Looking at it another way, we see 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) between two events can be the same because both the distance *and* the time between the events are different in the two frames.

Exercise 1.8. 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. (1.8), gives only the distance along a *straight line* between two points. The differential version of Eq. (1.8) gives the distance ds between *neighboring* points along *any* curve:

The Metric Postulate for a Flat Surface, Local Form

Let ds be the distance between neighboring points on a curve. Let the points have coordinate differences (dx, dy) in a planar frame. Then

$$ds^2 = dx^2 + dy^2. \tag{1.13}$$

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

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

The formula can give different lengths for different curves between two points.

Exercise 1.9. Use the formula to 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. (1.9) is concerned only with times measured by *inertial* clocks. The differential version of Eq. (1.9) gives the time ds measured by *any* clock between *neighboring* events on its worldline:

The Metric Postulate for a Flat Spacetime, Local Form

If a pulse of light can move between neighboring events, set $ds = 0$. If a clock can move between the events, let ds be the time it measures between them. Let the events have coordinate differences (dt, dx, dy, dz) in an inertial frame. Then

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2. \tag{1.14}$$

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

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

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 1.10. 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 1.11. 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 in an inertial frame I. Notation: Δt is the duration of the trip in the inertial frame, v_r is the velocity in I of the clock remaining on the ground, and Δs_r is the time the clock measures for the trip. Define v_a and Δs_a similarly for the airplane. Use Exercise 1.10 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 to obtain $\Delta s_a - \Delta s_r = 1.4 \times 10^{-7}$ s.

Exercise 2.1 shows that general relativity predicts a further difference between the clocks.

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. (1.11), 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/sec^2, and so this is a test of the local form Eq. (1.14) of the metric postulate as well as the original form Eq. (1.9).

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 *one* clock. Again by our definition of synchronized clocks, this two way speed is equal to the one way speed between P and 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. Their experiment is described in Appendix 7. It was repeated in 1930 by G. Joos, who found that any difference in the two way speeds is less than six parts in 10^{12}. A modern version of the experiment 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 7.

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 7. 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 it would 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.)

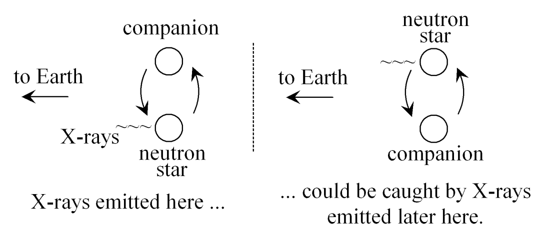


Fig. 1.10: The speed of light is independent of the speed of its source.

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

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.

1.4 The Geodesic Postulate

Flat Surface Geodesics. It is convenient to use superscripts to distinguish coordinates. Thus we use (x^1, x^2) instead of (x, y) for planar frame coordinates.

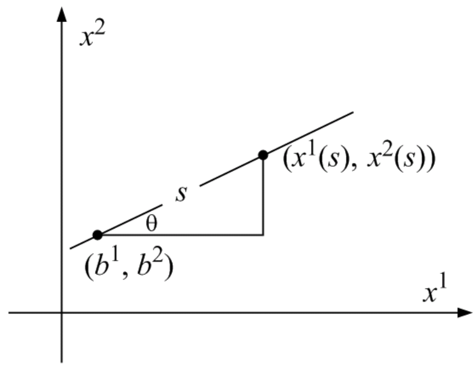


Fig. 1.11: A geodesic in a planar frame.

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

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

(The overdots indicate derivatives: $\ddot{x}^i(s) = d^2x^i/ds^2$.) The straight lines are called *geodesics*.

Not all parameterizations of a straight line satisfy the geodesic differential equations Eq. (1.15). For example, $x^i(p) = a_i p^3 + b_i$ parameterizes the same straight line as does $x^i(p) = a_i p + b_i$.

Flat Spacetime Geodesics. It is 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

$$x^i = a_i x^0 + b_i, \;\; i = 1, 2, 3. \tag{1.16}$$

(Differentiate to give $dx^i/dx^0 = a_i$; the velocity is constant.) The same assumption is made in prerelativity physics, where it is known as 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

$$x^i(p) = a_i p + b_i, \;\; i = 0, 1, 2, 3 \tag{1.17}$$

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 1.12. In Eq. (1.17) the parameter $p = x^0$. Show that the worldline of an inertial particle can also 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

$$\boxed{\ddot{x}^i(p) = 0, \;\; i = 0, 1, 2, 3.} \tag{1.18}$$

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

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

Exercise 1.13. Make as long a list as you can of analogous properties of flat surfaces and flat spacetimes.

Chapter 2

Curved Spacetimes

2.1 Newton's theory 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 flat surfaces and flat spacetimes 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.

Fig. 2.1: Isaac Newton, 1643-1727.

Before studying general relativity, we take a brief look at Isaac Newton's 1687 theory of gravity, the prevailing theory of gravity when Einstein formulated his theory.

Newton knew of the work of Johannes Kepler and Galileo Galilei from the early seventeenth century. Kepler discovered that the path of a planet is an ellipse with the Sun at one focus. Galileo discovered two important facts about objects falling near the Earth's surface: the acceleration is *constant* in time and *independent* of the mass and composition of the falling object.

Newton's theory explained Kepler's astronomical and Galileo's terrestrial findings as manifestations of the same phenomenon. To understand how orbital motion is related to falling motion, refer to Fig. 2.2. 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 attracted by the central body (do not ask how!), causing an acceleration

$$\boxed{a = -\frac{\kappa M}{r^2}\,,} \tag{2.1}$$

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.

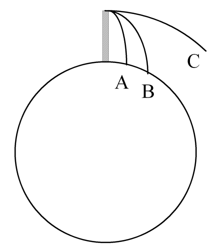


Fig. 2.2: Falling and orbital motion are the same.

Eq. (2.1) implies that the planets orbit the Sun in ellipses, in accord with Kepler's findings. See Appendix 8. By taking the distance r to the Earth's center to be essentially constant near the Earth's surface, 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.

Today we understand that the rest of the universe is made of the same "stuff" as here on Earth and obeys the same physical laws. Newton's theory of gravity was an early contribution to this great advance in our understanding of our place in the universe.

Newton's theory has enjoyed enormous success. Perhaps the most 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 astronomer, calculated that a new planet, beyond Uranus, could account for the discrepancy. He wrote J. Galle, an astronomer at the Berlin observatory, 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. In particular, 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 curved surfaces: *a small region of a curved surface is in many respects like a small region of a flat surface.* This is familiar: a small region of a (perfectly) spherical Earth appears 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 in many respects 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 in the airplane 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 Einstein's key observation: *as viewed by inertial observers*, a small region of a curved spacetime is in many respects 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. Newton's 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 close this section with a discussion of two kinds of experimental evidence for the postulates: universal acceleration and the gravitational redshift.

Universal Acceleration. Experiments of R. Dicke and of V. B. Braginsky, performed in the 1960's, verify to extraordinary accuracy one of Galileo's findings 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.

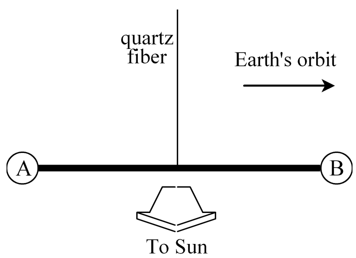


Fig. 2.3: Masses A and B accelerate the same toward the Sun.

Dicke and Braginsky used the Sun's gravity. The principle of their experiments can be seen in the simplified diagram in Fig. 2.3. 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 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 related experiment shows that the Earth and the Moon, despite the 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 the *lunar laser ranging experiment*. See Figs. 2.4 and 2.5.

Fig. 2.4: Lunar laser ranging experiment laser

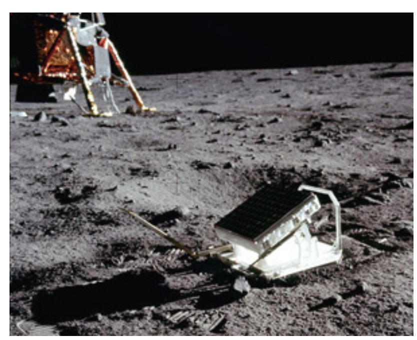


Fig. 2.5: Lunar laser ranging experiment mirrors on the Moon.

The measurements show that the Earth and Moon accelerate the same toward the Sun within one part in 10^4.

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. (2.28).

There is another difference between the Earth and the Moon. 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, only 5 parts in 10^{10}. But it is 25 times smaller for the Moon. One can wonder whether this difference between the Earth and 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.

Gravitational Redshift. 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. 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.

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 6 – 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

$$z = \frac{v}{c} = \frac{gh}{c^2}. \tag{2.2}$$

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 3.6 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 1.11. Show that the difference between the clocks due to the gravitational redshift is

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

Suppose $h = 10$ 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 1.11 gives

$$\Delta s_a - \Delta s_r = \left(\frac{1}{2}(v_a^2 - v_r^2) + gh\right)\Delta t = 3.0 \times 10^{-7} \text{ sec.}$$

Exercise 3.5 shows that a rigorous calculation in general relativity gives the same result.

In 2010 the gravitational redshift was verified to better than 1 part in 10^8 using quantum clocks. In the experiment h was about .1mm!

The satellites of the Global Positioning System (GPS) carry atomic clocks. A GPS receiver on Earth can determine its position within a few meters. Time dilation and gravitational redshifts must be taken into account for the system to function properly. If they were ignored, navigational errors of 10 km/day would accumulate.

2.3 The Local Inertial Frame Postulate

Local Planar Frames. Suppose that curved surface dwellers start to construct a square coordinate grid using identical 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. See Fig. 2.6. In smaller regions around P, the grid must become more square.

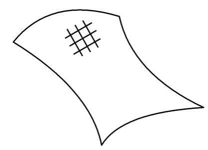


Fig. 2.6: A local planar frame.

The Local Planar Frame Postulate for a Curved Surface

A local planar frame can be constructed at any point P and with any orientation.

Global Surface Coordinates. We shall see that local planar frames at P provide surface dwellers with an intuitive description of many 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.

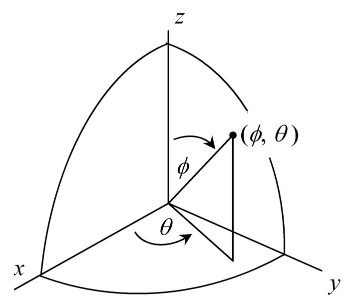


Fig. 2.7: Spherical coordinates (ϕ, θ) on a sphere.

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:

$$x = x(y^1, y^2),\ y = y(y^1, y^2),\ z = z(y^1, y^2)\,. \tag{2.3}$$

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. 2.7 shows spherical coordinates $(y^1, y^2) = (\phi, \theta)$ on a sphere of radius R. We have

$$x = R\sin\phi\cos\theta,\ y = R\sin\phi\sin\theta,\ z = R\cos\phi. \tag{2.4}$$

Exercise 2.2. Show that surface dwellers living on a sphere can assign (ϕ, θ) coordinates to their surface.

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.)

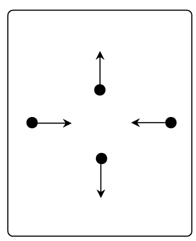


Fig. 2.8: Tidal accelerations in a radially free falling cabin.

Local Inertial Frames. 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. 2.8. 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.

The European Space Agency's GOCE satellite, planned for launch in 2009, is dedicated to measuring the Earth's gravity field in unprecedented detail. It will carry a gravity *gradiometer* to measure tidal accelerations. The gradiometer is essentially a sophisticated version of Figure 2.8. See Figure 2.9. As the satellite orbits, the tidal accelerations will vary due to mountains, changes in the density inside the Earth, etc. Among other benefits, the measurements will improve our understanding of the Earth's internal structure and provide a much better reference for ocean and climate studies, including sea-level changes and ice-sheet dynamics.

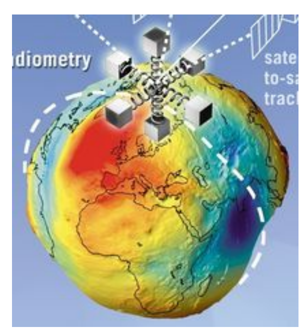


Fig. 2.9: The GOCE gradiometer.

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 inertial 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.

The Local Inertial Frame Postulate for a Curved Spacetime

Let E be an event on the worldline of an inertial object. A local inertial frame can be constructed at E with the inertial object at rest in it at E and with any given spatial orientation.

Global Spacetime Coordinates. We shall find that local inertial frames at E provide an intuitive description of many 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, a coordinate system is obtained by sending out a flash of light in all directions from an event. The flash is received at four airplanes flying on arbitrary paths. Each airplane has a clock. The clocks need neither to be synchronized nor to run uniformly. The arrival times of the flash are the coordinates of the event.

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

Curved Surface Metric. Recall the local form of the metric for flat surfaces, Eq. (1.13): $ds^2 = (dx^1)^2 + (dx^2)^2$.

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

$$ds^2 = (dx^1)^2 + (dx^2)^2. \tag{2.5}$$

See Fig. 2.6.

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

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

Then Eq. (2.5) can be written

$$ds^2 = \sum_{m,n=1}^{2} f^\circ_{mn} dx^m dx^n. \tag{2.6}$$

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

$$ds^2 = f^\circ_{mn} dx^m dx^n. \tag{2.7}$$

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

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. 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

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

where we have set

$$g_{jk}(y) = f^{\circ}_{mn}\,\frac{\partial x^m}{\partial y^j}\,\frac{\partial x^n}{\partial y^k}\,. \tag{2.9}$$

Since the matrix (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. Then there is a symmetric matrix $\mathbf{g}(y^i) = (g_{jk}(y^i))$ such that the distance ds from (y^i) to a neighboring point $(y^i + dy^i)$ is given by

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

The matrix $\mathbf{g}$ is called the *metric* of the surface (with respect to (y^i)).

Exercise 2.3. 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.,

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

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 $dx^2 = Rd\phi$. See Fig. 2.12.

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

Exercise 2.4. 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 learn in linear algebra that we should not think of a vector as a list of components (v_i), but as a single object $\mathbf{v}$ which represents a magnitude and direction (an arrow), independently of any coordinate system. If a coordinate system is introduced, then the vector acquires components. They will be different in different coordinate systems.

Similarly, we should not think of a metric as a list of components (g_{jk}), but as a single object $\mathbf{g}$ which represents infinitesimal distances, independently of any coordinate system. If a coordinate system is introduced the metric acquires components. They will be different in different coordinate systems.

Exercise 2.5. **a.** 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

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

Hint: See Eq. (2.8).

b. Show by direct calculation that the metrics of Exercises 2.3 and 2.4 are related by Eq. (2.12).

Curved Spacetime Metric. Recall the local form of the metric for flat spacetimes, Eq. (1.14): $ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$.

The Metric Postulate for a Curved Spacetime, Local Form

If a pulse of light can move between neighboring events, set $ds = 0$. If a clock can move between the events, let ds be the time it measures between them. Let F have coordinates (dx^i) in a local inertial frame at E. Then

$$\boxed{ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2.} \tag{2.13}$$

The metric postulate asserts *a universal light speed and a slowing of moving clocks in local inertial frames.* (See the discussion following Eq. (1.10).)

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

$$ds^2 = f^\circ_{mn} dx^m dx^n, \tag{2.14}$$

where

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

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. Then there is a symmetric matrix $\mathbf{g}(y^i) = (g_{jk}(y^i))$ such that the interval ds between (y^i) and a neighboring event $(y^i + dy^i)$ is given by

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

The matrix $\mathbf{g}$ is called the *metric* of the spacetime (with respect to (y^i)).

2.5 The Geodesic Postulate

Curved Surface Geodesics. Curved surface dwellers find that some curves in their surface are straight in local planar frames. They call these curves *geodesics*. Fig. 2.10 indicates that the equator is a geodesic but the other circles of latitude are not. A geodesic is as straight as possible, given that it is constrained to the surface. To traverse a geodesic, a surface dweller need only always walk "straight ahead" over hill and dale.

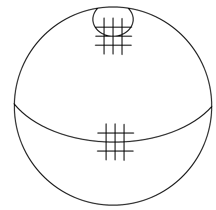


Fig. 2.10: The equator is the only circle of latitude which is a geodesic.

Recall the geodesic equations for flat surfaces, Eq. (1.15): $\ddot{x}^i = 0,\ i = 1, 2$.

The Geodesic Postulate for a Curved Surface, Local Form

Parameterize a geodesic with arclength s. Let point P be on the geodesic. Then in every local planar frame at P

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

We now translate Eq. (2.16) into 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 a matrix inverse $\mathbf{g}^{-1}$, which we denote (g^{jk}).

Exercise 2.6. 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 $\mathbf{g} = \mathbf{a}^t\mathbf{f}^\circ\mathbf{a}$, where t means "transpose".

c. Show 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*:

$$\Gamma^i_{jk} = \tfrac{1}{2}\, g^{im} \left[\partial_k g_{jm} + \partial_j g_{mk} - \partial_m g_{jk}\right]. \tag{2.17}$$

Note that $\Gamma^i_{jk} = \Gamma^i_{kj}$. The Γ^i_{jk}, like the g_{jk}, vary from point to point on a surface.

Exercise 2.7. Show that for the metric $\mathbf{g}$ of Eq. (2.11), $\Gamma^\phi_{\theta\theta} = -\sin\phi\,\cos\phi$. The only other nonzero Christoffel symbols are $\Gamma^\theta_{\theta\phi} = \Gamma^\theta_{\phi\theta} = \cot\phi$.

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

$$\ddot{y}^i + \Gamma^i_{jk}\dot{y}^j\dot{y}^k = 0, \quad i = 1, 2. \tag{2.18}$$

Appendix 10 translates the local form of the geodesic 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), $(f_{mn}(P)) = \mathbf{f}^\circ$. It is a mathematical fact that there are coordinates satisfying this relationship and also

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

for all m, n, i. They are called *geodesic coordinates*. See Appendix 9. 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 require that a local planar frame at P is a geodesic coordinate system at P.

Exercise 2.8. Show that the metric of Exercise 2.4 satisfies Eq. (2.19) at $(x, y) = (0, 0)$.

Exercise 2.9. Show that Eq. (2.18) reduces to Eq. (2.16) for local planar frames.

Exercise 2.10. 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 2.7. Before using the geodesic equations parameterize the circles with s.

Curved Spacetime Geodesics. Recall the geodesic equations for a flat spacetime, Eq. (1.18): $\ddot{x}^i = 0,\ i = 0, 1, 2, 3$.

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

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

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

The worldlines are called *geodesics*.

The geodesic postulate asserts that *inertial particles and light move in a straight line at constant speed in a local inertial frame.* (See the remarks after Eq. (1.18).) The worldline of an inertial particle or a light pulse 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. 2.11.

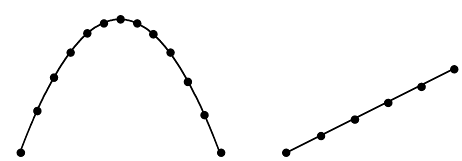


Fig. 2.11: The path of an inertial particle in a lattice stuck to the Earth and in a free falling inertial lattice. The dots are spaced at equal time intervals.

The postulate does not mention the mass or chemical composition of an inertial particle, so its worldline is independent of these attributes. Many sensitive experiments have confirmed this.

Einstein's "straightest worldline in a curved spacetime" description of the path is very different from Newton's "curved path in a flat space" description.

An analogy makes the difference more vivid. Imagine two surface dwellers living on a sphere. They start some distance apart on their equator and walk due north at the same speed along a circle of longitude (a geodesic). As they walk, they come closer together. They might attribute this to an attractive force between them. But we see that it is due to the curvature of their surface.

The local form of the geodesic postulate for curved spacetimes translates to

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

$$\boxed{\ddot{y}^i + \Gamma^i_{jk}\, \dot{y}^j \dot{y}^k = 0, \quad i = 0, 1, 2, 3.} \tag{2.21}$$

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

2.6 The Field Equation

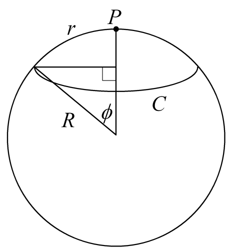


Fig. 2.12: On a sphere $K = 1/R^2$.

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.

Surface Curvature. The local forms of our curved surface postulates are examples of Gauss' key observation from Sec. 2.2: a small region of a curved surface is in many respects like a small region of a flat surface. Surface dwellers might suppose that *all* differences between such regions vanish as the regions become smaller. They would be wrong. To see this, pass geodesics through a point P in every direction.

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:

$$\boxed{K = \frac{3}{\pi} \lim_{r \to 0} \frac{2\pi\, r - C(r)}{r^3}.} \tag{2.22}$$

Clearly, $K = 0$ for a flat surface. From Fig. 2.12, $C(r) < 2\pi r$ for a sphere and so $K \geq 0$. From Fig. 2.12, we find

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

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. 2.13 is a *pseudosphere*. The horizontal "circles of latitude" are concave inward and the vertical "lines of longitude" are concave outward. Because of this "wiggling", $C(r) > 2\pi r$ 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.

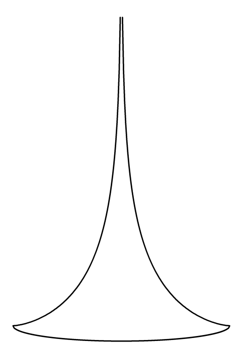


Fig. 2.13: On a pseudosphere $K = -1/R^2$.

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.

Roll a flat piece of paper into a cylinder. Since this does not distort distances on the paper, it does not change K. Thus $K = 0$ for the cylinder. Viewed from the outside, the cylinder is curved, and so $K = 0$ seems "wrong". However, surface dwellers restricted to a small region cannot determine that they now live on a cylinder. Thus K "should" be zero for a cylinder. (They can determine that they live on a cylinder by circumnavigating it.)

A formula expressing K in terms of distances was given by Gauss. In the special case $g_{12} = 0$ (setting $\partial_i \equiv \partial/\partial y^i$)

$$K = -\left(g_{11}g_{22}\right)^{-\frac{1}{2}}\left\{\partial_1\left(g_{11}^{-\frac{1}{2}}\,\partial_1 g_{22}^{\frac{1}{2}}\right) + \partial_2\left(g_{22}^{-\frac{1}{2}}\,\partial_2 g_{11}^{\frac{1}{2}}\right)\right\}. \tag{2.23}$$

Exercise 2.12. Show that Eq. (2.23) gives $K = 1/R^2$ for a sphere of radius R. Use the metric of a sphere, 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. 2.13. 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).

Spacetime Curvature. The local forms of our curved spacetime postulates are examples of Einstein's key observation from Sec. 2.2: a small region of a curved spacetime is in many respects like a small region of a flat spacetime. We might suppose that *all* differences between such regions vanish as the regions become smaller. We would be wrong.

To see this, refer to Fig. 2.8. Let $\Delta\, r$ be the small distance between objects at the top and bottom of the cabin and let $\Delta\, a$ 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 $\Delta\, a \to 0$ as $\Delta\, r \to 0$; 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: using Eq. (2.1), $\Delta\, a/\Delta\, r \to da/dr = 2\kappa M/r^3 \neq 0$. General relativity interprets this difference as a manifestation of the curvature of spacetime.

The Field Equation. The metric and geodesic postulates describe the behavior of clocks, light, and inertial particles in a curved spacetime. 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

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

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

$$R_{jk} = \Gamma^p_{tk}\,\Gamma^t_{jp} - \Gamma^p_{tp}\,\Gamma^t_{jk} + \partial_k \Gamma^p_{jp} - \partial_p \Gamma^p_{jk}\,. \tag{2.25}$$

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

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*

$$\mathbf{G} = \mathbf{R} - \tfrac{1}{2} R\, \mathbf{g}. \tag{2.26}$$

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 inertial matter; there are no pressures, stresses, or viscosity. Then the matter is called *dust*. Dust interacts only gravitationally. 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 independent of any coordinate system.) 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

$$\boxed{T^{jk} = \rho \frac{dy^j}{ds}\frac{dy^k}{ds}.} \tag{2.27}$$

We can now state our final postulate for general relativity:

The Field Equation

$$\boxed{\mathbf{G} = -8\pi\kappa\mathbf{T}.} \tag{2.28}$$

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

The field equation is the centerpiece of Einstein's theory. It relates the curvature of spacetime at an event, represented by $\mathbf{G}$, to the density and motion of matter and energy at the event, represented by $\mathbf{T}$. In our applications, we will specify $\mathbf{T}$, and then solve the equation for $\mathbf{G}$.

Tracing back through the definitions of $\mathbf{G}$, $\mathbf{R}$, and Γ^i_{jk} (Eqs. (2.26), (2.25), (2.17)) shows that $\mathbf{G}$, like the curvature K of a surface, involves the metric components g_{jk} and their first and second partial derivatives. Thus *the field equation is a system of second order partial differential equations in the unknown g_{jk}.*

Appendix 13 gives a plausibility argument, based on reasonable assumptions, which leads from the schematic field equation Eq. (2.24) to the field equation Eq. (2.28). It is by no means 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.

Covariant and contravariant. 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* from coordinates (y^i) to coordinates $(\bar{y}^p)$:

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

Exercise 2.16. Show that if a_{jk} and b_{jk} transform covariantly, then so does $a_{jk} + b_{jk}$.

We have seen the covariant transformation law Eq. (2.29) before, in Exercise 2.5, which shows that 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 $G_{jk} = R_{jk} - \frac{1}{2}Rg_{jk}$ transforms covariantly.

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

$$\bar{a}^{pq} = a^{jk}\,\frac{\partial \bar{y}^p}{\partial y^j}\,\frac{\partial \bar{y}^q}{\partial y^k}\,. \tag{2.30}$$

Exercise 2.17. Show that the T^{jk} transform contravariantly.

Exercise 2.18. Show that if a^j_k has one covariant and one contravariant index, then a^j_j is a *scalar*, i.e., it has the same value in all coordinate systems.

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, Eq. (2.28): 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 A.6 shows that a raised covariant index is indeed a contravariant index.

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 field equation $G_{jk} = -8\pi\kappa T_{jk}$.

[1] This 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 11 gives a short introduction to tensors.

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

$$0 = g^{jk}\left(R_{jk} - \tfrac{1}{2}R\,g_{jk}\right) = g^{jk}R_{jk} - \tfrac{1}{2}R\,g^{jk}g_{jk} = R - \tfrac{1}{2}R\,4 = -R\,.$$

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

The Vacuum Field Equation

$$\boxed{\mathbf{R} = \mathbf{0}\,.} \tag{2.31}$$

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

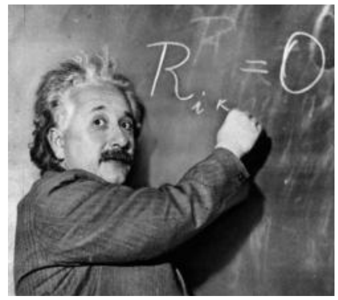


Fig. 2.14: Einstein's vacuum field equation.

The metric, the geodesic equations, and the field equation are examples of our ability to describe a curved spacetime intrinsically in arbitrary coordinates. You should be impressed with the power of this mathematics!

Curvature. The field equation yields a simple and elegant relationship between the density of matter at an event and the curvature of spacetime at the event. To obtain it we use Fermi normal coordinates, discussed in Appendix 12.

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 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 12 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 $\rho = 0$ at an event, then $K_{12} + K_{23} + K_{31} = 0$ at the event.

Chapter 3

Spherically Symmetric Spacetimes

3.1 Stellar Evolution

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

Interstellar gas and dust are major components of galaxies. Suppose a perturbing force causes a cloud of gas and dust to begin to contract by self gravitation. As the cloud contracts it will 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! See Fig. 3.1. Our star, the Sun, formed in this way 4.6 billion years ago.

Fig. 3.1: New stars shining within the cloud of gas and dust from which they formed.

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 $\lesssim 1.4$ solar masses, 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 10^6g/cm^3! Such a star is a *white dwarf*. They are common. For example, Sirius, the brightest star in the sky, is a member of a double star system. Its dim companion, Sirius B, is a white dwarf. See Fig. 3.2. The Sun will become a white dwarf in about 5 billion years.

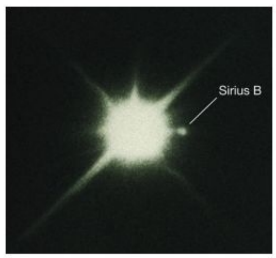


Fig. 3.2: Sirius and its white dwarf companion. The spikes on Sirius are an artifact.

Fig. 3.3: The crab nebula.

For a star over 1.4 solar masses, degenerate electron pressure cannot stop the contraction. Then a tremendous nuclear explosion, called a Type II *supernova*, occurs. The outer portions of the star are blown into interstellar space. A nearby supernova occurred in 1054. It was visible during the day for 23 days, outshining all other stars in the sky for several weeks. Today we see the material blown from the star as the *Crab Nebula*. See Fig. 3.3.

One possible result of a Type II supernova is 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 then a *pulsar*. The first pulsar was discovered in 1968. Many hundreds 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.

Degenerate neutron pressure cannot stop the contraction of a too massive star; the star will collapse to a *black hole*. Sec. 3.6 is devoted to black holes.

White dwarfs in binary star system are involved in *Type Ia* supernovae. At its brightest, a Type Ia supernova has a luminosity over a billion times that of the Sun. There are two ways a Type Ia supernova can occur. In the first, the white dwarf accretes matter from its companion. See Fig. 3.4. If it accretes sufficient matter, then it will erupt in a supernova, destroying the star. In the second, two orbiting white dwarfs spiral toward each other until they collide, producing the supernova.

Fig. 3.4: Artist's depiction of a white dwarf accreting matter from a companion.

We close our catalog of remarkable astronomical objects with *quasars*, discovered in 1963. Quasars sit at the center of some distant galaxies. Their outstanding feature is their enormous energy output, typically 100 times that of our entire Milky Way galaxy from a region 10^{17} times smaller!

3.2 Schwarzschild Metric

In this section we investigate the metric of a spacetime around an object which is spherically symmetric and unchanging in time, for example, the Sun. This will enable us in the next section to compare general relativity's predictions for the motion of planets and light in our solar system with observations.

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

$$ds^2 = dt^2 - dr^2 - r^2 d\Omega^2, \tag{3.1}$$

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 retain their usual meaning.
- $d\theta$ and $-d\theta$ produce the same ds, as do $d\phi$ and $-d\phi$. Thus none of the terms $drd\theta, drd\phi, d\theta d\phi, dtd\theta,$ or $dtd\phi$ 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

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

where $\mu, \nu, u,$ and λ are unknown functions of r. They are not functions of θ or ϕ (by spherical symmetry) or of t (since the metric is unchanging in time). We write some of the coefficients as exponentials for convenience.

Exercise 3.2. Show that the coordinate change $\bar{r} = re^{\lambda(r)}$ eliminates the $e^{2\lambda}$ factor. (Then name the radial coordinate r again.)

Exercise 3.3. Show that the coordinate change $\bar{t} = t + \Phi(r)$, where $\Phi'(r) = -u\mathrm{e}^{-2\mu}$, eliminates the $dtdr$ term. (Then name the time coordinate t again.)

These coordinate changes put the metric in a simpler form:

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

This is as far as we can go with spherical symmetry and coordinate changes. How can we determine the unknowns μ and ν in the metric? Well, the field equation determines the metric. And since we are interested only in the spacetime outside the central object, we may use the vacuum field equation $\mathbf{R} = \mathbf{0}$.

In components the vacuum field equation says $R_{jk} = 0$. The R_{jk} are best calculated with a computer. We need these three (a prime indicates differentiation with respect to r):

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

Set $R_{tt} = 0$ and $R_{rr} = 0$, cancel the exponential factor, and subtract: $\mu' + \nu' = 0$. Thus $\mu + \nu = C$, a constant. Now $R_{\phi\phi} = 0$ implies $2\,r\nu' + \mathrm{e}^{2\nu} - 1 = 0$, or $(r\,\mathrm{e}^{-2\nu})' = 1$. Integrate: $\mathrm{e}^{-2\nu} = 1 - 2\,m/r$, where $2\,m$ is a constant of integration. The metric Eq. (3.3) can now be put in the form

$$ds^2 = \left(1 - \frac{2m}{r}\right) \mathrm{e}^{2C} dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2 d\,\Omega^2.$$

The substitution $d\bar{t} = \mathrm{e}^{C} dt$ eliminates the e^{2C} factor. We arrive at our solution, the *Schwarzschild metric*, obtained by Karl Schwarzschild in 1916:

Schwarzschild Metric

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

In matrix form the metric is

$$\mathbf{g} = (g_{jk}) = \begin{pmatrix} 1 - \frac{2m}{r} & 0 & 0 & 0 \\ 0 & -\left(1 - \frac{2m}{r}\right)^{-1} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2 \sin^2\phi \end{pmatrix}.$$

We shall see later that

$$m = \kappa M, \tag{3.6}$$

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

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

If $r \gg 2m$, then the Schwarzschild 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 3.4. Show that the time ds measured by a clock at rest at r is related to the coordinate time dt by

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

Exercise 3.5. Show that the Schwarzschild metric predicts the sum of the results of Exercises 1.11 (time dilation) and 2.1 (gravitational redshift) for the 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. Eq. (2.2) gives a formula for the gravitational redshift over short distances. Here we generalize to arbitrary distances using the Schwarzschild metric. Emit pulses of light radially outward from (t_e, r_e) and $(t_e + \Delta t_e, r_e)$ in a Schwarzschild spacetime. Observe them at (t_o, r_o) and $(t_o + \Delta t_o, r_o)$. Since the Schwarzschild metric is time independent, the worldline of the second pulse is simply a time 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 rest at r_e, with Δs_o defined similarly. By Eqs. (1.6) and (3.7) an observer at r_o finds a redshift

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

Exercise 3.6. 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 3.7. 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.

The most 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 Doppler 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) for the spacetime. They are:

$$\ddot{t} + \frac{2m/r^2}{1-2m/r}\,\dot{r}\,\dot{t} = 0 \tag{3.9}$$

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

$$\ddot{\theta} + \frac{2}{r}\,\dot{r}\,\dot{\theta} + 2\cot\!\phi\,\dot{\phi}\,\dot{\theta} = 0 \tag{3.11}$$

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

Exercise 3.8. What are the Christoffel symbols $\Gamma^{\,r}_{tt}$ and $\Gamma^{\,t}_{tr}$?

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

$$\phi = \pi/2\,. \tag{3.13}$$

This is a solution to Eq. (3.12).

Exercise 3.9. 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 3.4.) This shows that a particle (with $ds > 0$) can have a circular orbit only for $r > 3m$ and 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):

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

$$\dot{\theta}\,r^2 = A, \tag{3.15}$$

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 3.10. 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:

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

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 $E = 1$.

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.

Summary. It was a long journey to obtain the partially integrated geodesic equations, so it might be useful to outline the steps involved:

- We obtained a general form for the metric of a spherically symmetric time independent spacetime, Eq. (3.2), using coordinate changes and symmetry arguments.
- We wrote some of the components of the Ricci tensor for the spherically symmetric metric, Eqs. (3.4). They involve second derivatives of the metric.
- We wrote the vacuum field equations for the metric by setting the components of the Ricci tensor to zero.
- We obtained the Schwarzschild metric, Eq. (3.5), by solving the vacuum field equations.
- We wrote the geodesic equations for the Schwarzschild metric, Eqs. (3.9)–(3.12). They involve first derivatives of the metric and are of second order.
- We obtained the first order partially integrated geodesic equations, Eqs. (3.13)–(3.16), by integrating the second order geodesic equations.

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

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

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

$$\frac{d^2r}{ds^2} = -\frac{m}{r^2}. \tag{3.18}$$

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 Solar System Tests

This section discusses general relativity's predictions of the motion of planets and light in our solar system. There are minuscule differences from the predictions of Newton's theory. We shall describe three: perihelion advance, light deflection, and light delay. The predictions provide *tests* of general relativity: do measurements confirm them?

Perihelion Advance. To determine the motion of the planets, we must solve the geodesic equations. Set $u = 1/r$. Using 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 $2A^2 du/d\theta$:

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

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

Exercise 3.11. The ratio of the terms on the right side of Eq. (3.19) is $3mu^2/mA^{-2}$. Among the planets, Mercury has the largest ratio. Show that it is less than 10^{-7}.

Instead of solving Eq. (3.19) directly and exactly, we use a *method of successive approximations* to solve it approximately. As a first approximation, solve Eq. (3.19) without the small $3mu^2$ term:

$$u = mA^{-2}\left\{1 + e\,\cos(\theta - \theta_p)\right\}. \tag{3.20}$$

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 is derived from Newton's theory in Appendix 8. 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.

We now obtain a better approximation to the solution of Eq. (3.19). Set $\theta = \theta_p$ in Eq. (3.20):

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

Substitute Eqs. (3.20) and (3.21) into the right side of Eq. (3.19) and solve:

$$\begin{aligned} u = {} & r_p^{-1}\,(1+e)^{-1}\left\{1 + e\left[\cos(\theta - \theta_p) + 3mr_p^{-1}\,(1+e)^{-1}\,\theta\,\sin(\theta - \theta_p)\right]\right\} \\ & + mr_p^{-2}\,(1+e)^{-2}\left\{3 + \tfrac{1}{2}e^2\left[3 - \cos(2(\theta - \theta_p))\right]\right\}. \end{aligned} \tag{3.22}$$

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):

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

This is a good enough approximation to the solution of Eq. (3.19) for our purposes. To understand it, compare it to the equation of an an ellipse with perihelion θ_p, Eq. (3.20): $u = r_p^{-1}(1+e)^{-1}\{1 + e\,\cos(\theta - \theta_p)\}$ (using Eq. (3.21)). The equations are the same except for the term $3m\theta/r_p(1+e)$. Since $3m/r_p(1+e)$ is very small, $\theta_p + 3m\theta/r_p(1+e)$ in Eq. (3.23) is nearly constant over one revolution, $\theta \to \theta + 2\pi$. Thus over the revolution, Eq. (3.23) is nearly the equation of an ellipse. However, after the revolution the perihelion has *advanced* by $6m\pi/r_p(1+e)$. So our picture of the orbit is a slowly rotating ellipse. **Exercise 3.12**. Show that for Mercury the predicted perihelion advance is 43.0 arcsecond/century.

The perihelion of Mercury advances 574 arcsec/century. In 1845 Le Verrier showed that Newton's theory explains the advance as small gravitational effects of other planets – except for about 35 arcsec, for which he could not account. This was refined to 43 arcsec by the time Einstein published the general theory of relativity. It was the only discrepancy between Newton's theory and observation known at that time. Less than an arc*minute* 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. 3.5. 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

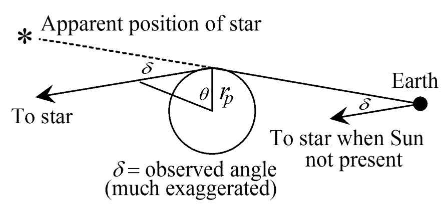


Fig. 3.5: Deflection of light by the sun.

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

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:

$$u = u_p\cos\theta + \tfrac{1}{2}mu_p^2(3 - \cos 2\theta). \tag{3.25}$$

At the star, $u = 0$ and $\theta = \pi/2 + \delta/2$. Substitute these into Eq. (3.25). Use

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

for small δ to obtain the deflection angle $\delta = 4m/r_p$. For the Sun, $\delta = 1.75$ arcsec. The Newtonian equation Eq. (2.1) predicts half of this deflection.

Stars can be seen near the Sun only during a solar eclipse. The prediction of general relativity was confirmed within about 20% during an eclipse in 1919. Later eclipse observations 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 in daylight.

A spectacular example of the gravitational deflection of light was discovered in 1979. Two quasars, 6 arcsec apart in the sky, are in fact the same quasar! Fig. 3.6 shows how an intervening galaxy causes a double image of the quasar. The galaxy is a *gravitational lens*. Many are now known. In one system, brightness variations in the quasar are seen in one image two years before the other. Why? Because the lengths of the two paths are different.

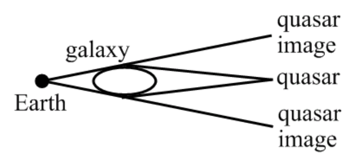


Fig. 3.6: A gravitational lens.

It is not necessary that the lensed galaxy be exactly centered behind the lensing galaxy. But in a few known cases this is nearly so. Then the lensed galaxy appears as a ring, called an *Einstein ring*, around the lensing galaxy! See Fig. 3.7.

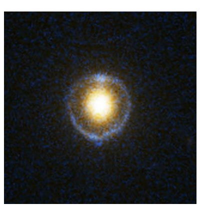


Fig. 3.7: An Einstein ring.

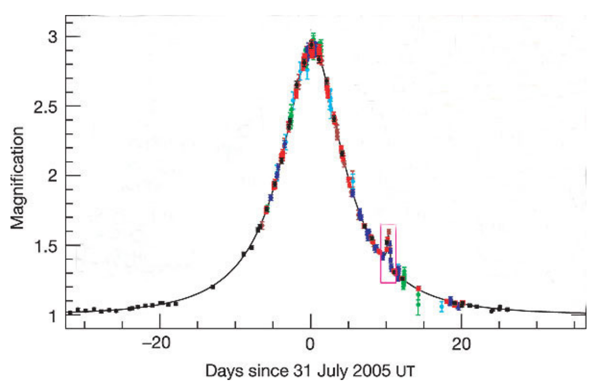


Fig. 3.8: Gravitational lensing by a star (main curve) and an accompanying planet (in the red rectangular box).

Gravitational lensing has been used to detect planets orbiting other stars. If a star passes in front of another, gravitational lensing by the foreground star will brighten the background star. Fig. 3.8 shows a star brightening to a maximum and then returning to normal. A planet superposes a small blip on the brightening.

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 its passage through the Sun's gravity. See Fig. 3.9.

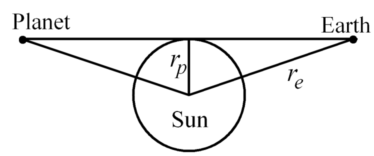


Fig. 3.9: Radar echo delay.

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 $r = r_s$ at the closest approach to the Sun. Then $\dot{r}|_{r=r_s} = 0$. Substitute into Eq. (3.16):

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

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

$$t = \int_{r_s}^{r_e} \left(\frac{\left(1 - \frac{2m}{r}\right)^{-2}}{1 - \frac{1-2m/r}{1-2m/r_s}\left(\frac{r_s}{r}\right)^2} \right)^{\frac{1}{2}} dr. \tag{3.26}$$

In Appendix 14 we approximate the integral:

$$t = (r_e^2 - r_s^2)^{\frac{1}{2}} + 2m \ln(2r_e/r_s) + m. \tag{3.27}$$

The first term, $(r_e^2 - r_s^2)^{\frac{1}{2}}$, 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. There are many difficulties in comparing this calculation with measurements. But they have been overcome and the prediction verified within 5%. Signals from the Cassini spacecraft to Saturn have confirmed this effect to better that one part in 10^4.

Quasars can vary in brightness on a timescale of months. We see a variation in the southern image of the original double quasar 417 days after the same variation of the northern image. The southern image is 1 arcsec from the lensing galaxy; that of the northern is 5 arcsec. Thus the light of the northern image travels farther 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, called *geodetic precession*, is motivated in Figs. 3.10 and 3.11. In Fig. 3.10, 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. 3.11 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.

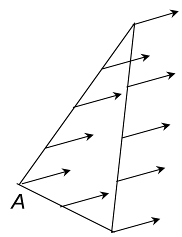


Fig. 3.10: A parallel transported vector returns to A with its original direction.

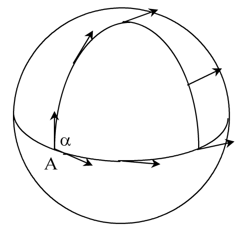


Fig. 3.11: A parallel transported vector returns to A rotated through angle α.

In a flat spacetime the axis of rotation of an inertial gyroscope moves parallel to itself in an inertial frame. The same is true in a curved spacetime in local inertial frames. 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 rotating Earth-Moon system is a "gyroscope" orbiting the Sun. The predicted geodetic precession is $\sim .02$ arcsec/yr. The lunar laser ranging experiment has confirmed this to .1%.

3.4 Kerr Spacetimes

The Kerr Metric. A rotating object has an axis of rotation, and so is not spherically symmetric. The spacetime around it is not described by the Schwarzschild metric. The *Kerr metric* is an axially symmetric solution of the field equations. It was discovered only in 1963 by Roy Kerr. At sufficiently large distances from a rotating object, the metric is close to the Kerr metric.

Kerr Metric

$$\begin{aligned} ds^2 = &\left(1 - \frac{2mr}{\Sigma}\right) dt^2 + \frac{4mar\sin^2\phi}{\Sigma}\, dt\, d\theta - \frac{\Sigma}{\Delta}\, dr^2 \\ &- \Sigma\, d\phi^2 - \left(r^2 + a^2 + \frac{2ma^2 r \sin^2\phi}{\Sigma}\right) \sin^2\phi\, d\theta^2, \end{aligned} \tag{3.28}$$

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 z-axis is the rotation axis.

Exercise 3.13. Show that if the central object is not rotating, then the Kerr metric Eq. (3.28) reduces to the Schwarzschild metric Eq. (3.5).

Note that the surfaces $t = \text{const}$, $r = \text{const}$ in the Kerr metric 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 a circular polar orbit at two Earth radii will change by $\sim .03$ arcsec/yr. This is a shift of ~ 2 m/yr in the intersection of the orbital and equatorial planes.

The *Gravity Probe B* experiment, launched in polar orbit in April 2004, was designed to accurately measure frame dragging on gyroscopes. The experiment is a technological tour-de-force. Its four 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 and the solar wind must not affect their 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. 3.12. Small rockets on the shell

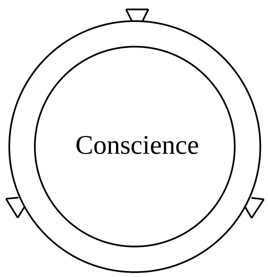


Fig. 3.12: The conscience.

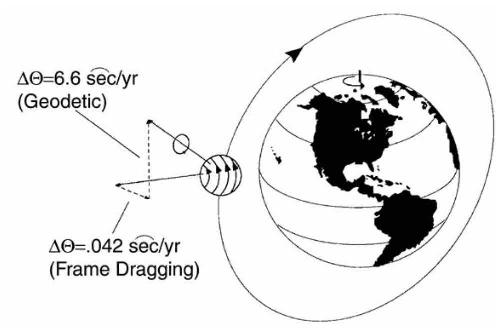


Fig. 3.13: The Gravity Probe B experiment.

keep it centered around the conscience when the shell deviates from inertial motion. Clever!

The predicted frame dragging precession is .04 arcsec/year. See Fig. 3.13. The experiment was designed to test this to about 1%, but only achieved, it is claimed, 20% accuracy. The experiment also measured geodetic precession. The predicted geodetic precession was 6.6 arcsec/year. The experiment was designed to test this to one part in .01%, but the final analysis of the data only achieved .3% accuracy.

Frame dragging has been measured to an accuracy of 5% by tracking the orbits of several completely passive laser ranged satellites in very stable Earth orbits. Each satellite is a small (< 1 meter) sphere.

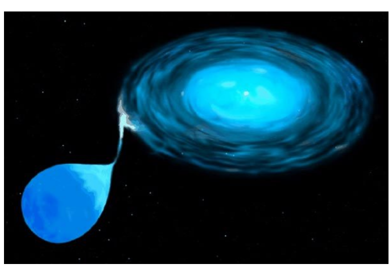


Fig. 3.14: Artist's depiction of an accretion disk.

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. (Sec. 3.6 discusses black holes.) Accretion disks form when a compact star (white dwarf, neutron star, or black hole) in a binary star system attracts matter from its companion, forming a rapidly rotating disk of hot gas around it. Fig. 3.14 shows an artist's depiction of an accretion disk. There is evidence that frame dragging causes the axis of some disks around black holes to precess hundreds of revolutions a second.

Clock effect. 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 $\theta \to \theta \pm 2\pi$. (After a complete orbit the satellite is not above the same point of the spinning central object.) We derive this *gravitomagnetic clock effect* from the t-geodesic equation for circular equatorial orbits in the Kerr metric:

$$\dot{t}^{\,2} - 2a\dot{t}\dot{\theta} + \left(a^2 - \frac{r^3}{m}\right)\dot{\theta}^2 = 0. \tag{3.29}$$

Solve for $dt/d\theta$:

$$\frac{dt}{d\theta} = \pm\left(\frac{r^3}{m}\right)^{\frac{1}{2}} + a. \tag{3.30}$$

If $a = 0$, then this reduces to the Schwarzschild value $dt/d\theta = \pm(r^3/m)^{\frac{1}{2}}$.

Consider first the orbit $\theta \to \theta + 2\pi$, with period Δt_+. Integrating Eq. (3.30) over the orbit multiplies the right side by 2π: $\Delta t_+ = 2\pi(r^3/m)^{\frac{1}{2}} + 2\pi a$. (Since $dt/d\theta > 0$, we use the "+" sign on the right side of Eq. (3.30).) Now consider the orbit $\theta \to \theta - 2\pi$. Integrating Eq. (3.30) over the orbit multiplies the right side by -2π: $\Delta t_- = 2\pi(r^3/m)^{\frac{1}{2}} - 2\pi a$. (Since $dt/d\theta < 0$, we use the "$-$" sign on the right side of Eq. (3.30).) Thus $\Delta t_+ - \Delta t_- = 4\pi a$; an orbit in the sense of a takes longer than its opposite orbit. The difference is independent of the orbital period.

Exercise 3.14. Show that for the Earth, $\Delta t_+ - \Delta t_- = 2 \times 10^{-7}$ sec. This has not been measured.

3.5 Extreme Binary Systems

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 stronger gravity and/or larger velocities. Both are present in some binary star systems consisting of two collapsed stars.

One spectacular system consisting of two pulsars was discovered in 2003. The system, called the *double pulsar*, provides the most stringent tests of general relativity to date. Fig. 3.15 shows an artist's depiction.

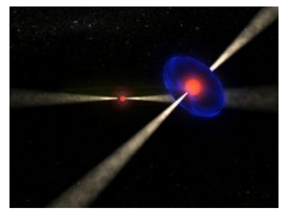


Fig. 3.15: Double pulsar. (Artist's depiction.)

One pulsar, of 1.4 solar masses, spins at 44 revolutions/second, the other, of 1.25 solar masses. at .36 revolutions/second. The pulsars are less than a solar diameter apart, with an orbital period of 2.4 hours. Imagine!

The pulsars' pulses serve as the ticking of very accurate clocks in the system. Several relativistic effects have been measured by analyzing the arrival times of the pulses at Earth. Several changing relativistic effects affect the arrival times: a gravitational redshift of the rate of clocks on Earth as the Earth's distance from the Sun changes over a year, a gravitational redshift of the time between the pulses as the pulsars move in their elliptical orbits, a time dilation of the pulsars' clocks due to their motion, and a light delay of the pulses. The measured delay is within .05% of the predictions of general relativity.

The periastron advance of the pulsars is 17°/year, 150,000 times that of Mercury. (Perihelion refers to the Sun; periastron is the generic term.) The measured geodetic precession in the system is 4.8°/year, within 13% of the predictions of general relativity. This is 5 million times larger than the precession in the Earth-Moon system! The accuracy will improve as observations extend over the years.

Most important, the system provides evidence for the existence of gravitational waves, which are 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. Violent astronomical phenomena, such as supernovae and the big bang, emit gravitational waves. So far, none has been directly detected on Earth. (But see Sec. 4.5.)

General relativity predicts that the orbital period of the double pulsar decreases by about 4×10^{-5} sec/year due to energy loss from the system from gravitational radiation. This is confirmed within 0.5%. The prediction is based on the full field equation, rather than the vacuum field equation used earlier in this chapter. Binary neutron stars provide the only precision test of the full field equation to date.

As gravitational waves carry energy away from the binary pulsar system, the stars come closer together. They will collide in about 85 million years, with an enormous burst of gravitational radiation. Such collisions are the most promising candidates for sources of gravitational waves detectable on Earth.

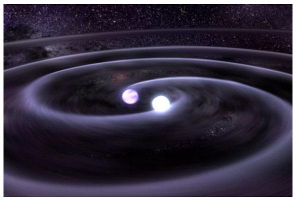


Fig. 3.16: Two white dwarfs with gravitational wave ripples in spacetime. (Artist's conception.)

Another system, discovered in 2011, consists of two white dwarfs in a 12.75 minute orbit. See Fig. 3.16. The system, designated J0651, is 3000 light years away. One of the stars has half of the Sun's mass and is about the size of Neptune. The other has a quarter of the Sun's mass and is about the size of the Earth. The stars eclipse each other during their orbit. Due to the emission of gravitational waves the eclipses occur five seconds (!) earlier over the course of a year. "This is a general relativistic effect you could measure with a wrist watch", said one researcher. The stars will merge in two million years.

3.6 Black Holes

Schwarzschild Black Holes. The $-(1 - 2m/r)^{-1}dr^2$ term in the Schwarzschild metric Eq. (3.5) has a singularity at $r = 2m$, the *Schwarzschild radius*. 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 relevance for the Sun or Earth.

However, we may inquire 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.

Suppose a material particle or pulse of light moves outward (not necessarily radially) from the horizon $r_1 = 2m$ 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 for the trip 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 reach r_2 at any finite t. Thus 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 reach the horizon from r_2. Thus a distant observer will not see the matter or light cross the horizon.

Exercise 3.15. Suppose an inertial object falling radially toward a black hole is close to the horizon, 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 gravitational redshift of an object approaching a Schwarzschild black hole increases to infinity as the object approaches the horizon.[1] A distant stationary observer will see the rate of all physical processes on the object slow to zero as it approaches the horizon. In particular, its brightness (rate of emission of light) will dim to zero and it will effectively disappear. But the observer will never see it cross the horizon.

On the other hand, an inertial observer radially approaching a Schwarzschild black hole will measure a finite time to cross the horizon. To see this, recall that a clock carried by the inertial observer measures proper time ds. By Eq. (3.17), $|dr/ds|$ increases as r decreases. Thus the observer will cross the horizon, never to return, in a finite proper time measured by the clock.

[1] Equation (3.8) applies to objects at rest. For objects falling toward the horizon, a Doppler effect causes an even larger redshift.

The Horizon. We can give the point with coordinates (x, y) in a plane new coordinates $(1/x, 1/y)$. Thus the point with usual coordinates $(2, 3)$ has new coordinates $(\frac{1}{2}, \frac{1}{3})$. The new coordinates provide a perfectly valid, if inconvenient, coordinate system – except when $x = 0$ or $y = 0$, where they have singularities. The problem is with the new coordinates; there is nothing wrong with the plane when $x = 0$ or $y = 0$.

In a similar way, the singularity at the horizon in the Schwarzschild metric is a *coordinate* singularity; there is no *physical* singularity in the spacetime. Someone crossing it would not notice anything special.

To see this, introduce *Painlevé*[2] *coordinates* with the coordinate change

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

in the Schwarzschild metric.

Exercise 3.16. Show that the metric in Painlevé coordinates is

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

The only singularity in this metric is at $r = 0$. Thus there is no physical singularity at the horizon. The $r = 0$ singularity is a physical singularity.

Kerr Black Holes. The spacetime of a Kerr black hole is even more peculiar than that of a Schwarzschild black hole. The $-(\Sigma/\Delta)dr^2$ term in the Kerr metric Eq. (3.28) has a singularity when $\Delta = 0$, the *horizon* of a Kerr black hole.

Exercise 3.17. Show that the horizon is at $r = m + (m^2 - a^2)^{\frac{1}{2}}$.

As with a Schwarzschild black hole, there is no singularity in the spacetime at the horizon. And 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.

A second surface of interest for a Kerr black hole is where $(1 - 2mr/\Sigma)\, dt^2 = 0$, the *static limit*. There is no singularity in the spacetime on this surface.

Exercise 3.18. Show that the static limit is at $r = m + (m^2 - a^2 \cos^2\phi)^{\frac{1}{2}}$.

Fig. 3.17 shows the axis of rotation of a Kerr black hole, the static limit, and the horizon. 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. (3.28) are negative. Since $ds^2 \geq 0$ for both particles and light, the $dtd\theta$ 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.

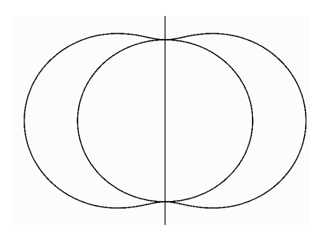


Fig. 3.17: The horizon (inner surface) and static limit (outer surface) of a Kerr black hole.

[2] Paul Painlevé, 1863-1933, was briefly Prime Minister of France.

Black Holes in Nature. General relativity allows black holes, but do they actually exist? The answer is "yes". They cannot be observed directly, as they emit no radiation. They must be inferred from their gravitational effects. 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. There have also been a few (only) detections of $10^2 - 10^4$ solar mass black holes.

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 star larger than $\sim$3 solar masses when it exhausts its nuclear fuel. Stellar masses range up to 100 solar masses. A massive star can shed a large fraction of its mass late in its life. Unless it ends up under three solar masses, it will become a black hole.

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. 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, and is therefore believed to be a black hole. In one case the companion has 20 solar masses.

During "normal" periods, gas pulled from the normal star and heated violently on its way to the compact companion causes the X-rays. Many 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 – topping off at at least 2×10^{10} solar masses! Gas can fall into the black hole in such large quantities that the gas is heated to high temperatures and emits vast quantities of electromagnetic radiation. This is the explanation of quasars.

There is a supermassive black hole at the center of our galaxy, about 26,000 light years away in the direction of the constellation Sagittarius. One star orbiting the black hole has eccentricity .87, closest approach to the hole 17 light hours while traveling 5000 km/sec, and period 15 years. This implies that the black hole is ≈ 3.5 million solar masses.

Exercise 3.19. Compute the Schwarzschild radius of this black hole and compare it to the radius of Mercury's orbit.

There is also good evidence for a small population of *intermediate* black holes, i.e., with mass between those of solar mass and supermassive black holes. Examples include 400, 5000, and $20,000$ solar mass black holes.

Most stars rotate. Due to conservation of angular momentum, a rotating star collapsed to a black hole will spin rapidly. Also, angular momentum can be transferred from an accretion disk to its black hole. An important theorem states that an (uncharged) rotating black hole will very quickly settle down to a Kerr black hole by emitting gravitational waves. For example, a collision of two black holes will not immediately produce a Kerr black hole, but it will quickly become one. Thus the Kerr metric is of fundamental astrophysical importance.

It is believed that here is a limit to the angular momentum of a Kerr black hole: the angular momentum parameter $a = J/M$ in the Kerr metric, Eq. (3.28), satisfies $|a/m| < 1$. The black hole at the center of our galaxy is spinning rapidly: there is good evidence that that $|a/m| \geq .6$.

A double supermassive black hole system has recently been analyzed. The larger black hole has 17 billion solar masses – the largest known. The other, of 100 million solar masses, orbits it at about 10% of the speed of light and with a period of 12 years. It crosses the accretion disk of the larger black hole twice each orbit, about a year apart, causing flares seen from Earth. See Figure 3.18. The timing of the flares was used to determine the orbit. It precesses 39°/revolution. The system is the "brightest" known source of gravitational radiation – 10^{10} brighter than the binary pulsar. The radiation causes the orbital period to decay by 20 days/orbit. The black holes will merge in less than ten thousand years.

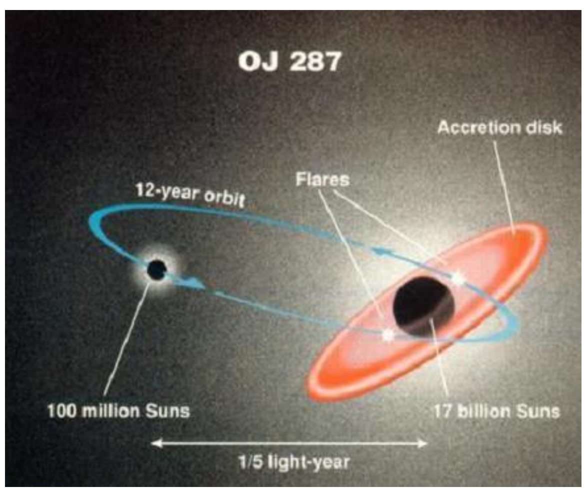


Fig. 3.18: A double supermassive black hole

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, from its origin, to today, and into the future. The universe is viewed as a single physical system. This is an audacious move, but we shall see that the model is very successful. We will begin with a description of the universe as seen from Earth.

I emphasize the word "seen" above. In Section 4.4, we will learn that most of the matter and energy in the universe is not seen as electromagnetic radiation, but is detected by other means.

First I note the following. There are lots of big numbers in this chapter. You need to get a sense of the relative size of these numbers; just thinking of them as "big" is not good enough. For example, a billion light years is to a million light years as a kilometer is to a meter.

Galaxies. The Sun is but one star of hundreds of billions held together gravitationally to form the *Milky Way galaxy.* The Milky Way is about 150,000 light years across. It is but one of tens of billions of galaxies in the visible universe. The *Andromeda galaxy*, about 2.5 million light years away, is the closest major galaxy to our own. See Fig. 4.1. Nearly everything you see in Fig. 4.2 is a galaxy. Importantly for us, galaxies are inertial objects.

Galaxies cluster in groups of up to several thousand. These are the largest structures in the unverse bound by gravity.

Fig. 4.1: The Andromeda galaxy.

Fig. 4.2: Galaxies galore!

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

$$\boxed{v = Hd\,.} \tag{4.1}$$

This is the "expansion of the universe". Eq. (4.1) is *Hubble's law*; H is *Hubble's constant.* Its value is 19.1 (km/sec)/million light years, within 1%.

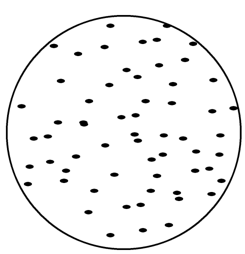


Fig. 4.3: Balloon analogy to the universe.

An expanding spherical balloon, on which bits of paper are glued, each representing a galaxy, provides a very instructive analogy. See Fig. 4.3. The balloon's two dimensional surface is the analog of our universe's three dimensional space. The galaxies are glued, not painted on, because they do not expand with the universe. From the viewpoint of *every* galaxy on the balloon, other galaxies recede from it, and Eq. (4.1) holds. Thus the the fact that galaxies recede from us does not mean that we occupy a special place in 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 surface 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.

What is the evidence for Hubble's law, $v = Hd$? We cannot verify it 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 $z > 0$. This is a key cosmological observation. 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, 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 13.8 billion years ago.

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.

Because the speed of light is finite, 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! Fig. 4.4 illustrates this. Astronomers study the early universe by studying the distant universe. The light from the farthest (oldest) known galaxy (so far) was emitted when the universe was only ~ 400 million years old. It has a redshift $z = 11.1$.

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

The Cosmic Microwave Background. 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, quite by accident and unaware of the prediction, discovered microwave radiation with the proper characteristics. This *cosmic microwave background* (CMB) is strong evidence that a big bang actually occurred.

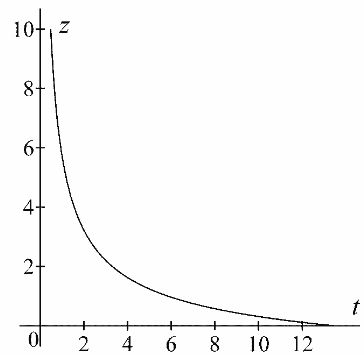


Fig. 4.4: The redshift z of a galaxy vs. the age of the universe t (in 10^9 years) when the light we see from it was emitted.

The CMB has a 2.7 K blackbody spectrum. The radiation started its journey toward us $\sim 370,000$ 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.

To a first approximation the CMB is isotropic, i.e., spherically symmetric about us. However, in 1977 astronomers measured a small dipole anisotropy. After correcting for the Earth's velocity around the Sun, the anisotropy is a relative redshift in the CMB of $z = -.0012 \cos\theta$, where θ is the angle from the constellation Leo.

Exercise 4.1. Explain this as due to the solar system moving toward Leo at 360 km/sec with respect to an isotropic CMB. In this sense we have discovered the absolute motion of the Earth.

After correcting for the dipole anisotropy, the CMB is isotropic to 1 part in 10^4.

Further evidence for a big bang comes from the theory of *big bang nucleosynthesis.* The theory predicts that in the first 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 this. 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.* Galaxies, stars, planets, life, and cultures all evolve.

4.2 The Robertson-Walker Metric

Our cosmological model assumes that the universe is exactly isotropic about every point. Most astronomers believe that that observations confirm this, but it has recently been questioned by some. The assumption spreads the galaxies into a uniform distribution of inertial matter, i.e., dust. This is rather like taking the Earth to be exactly spherical; it is not, but a sphere is an excellent approximation for many purposes.

We first set up a coordinate system (t, r, ϕ, θ) for the universe. As the universe expands, the temperature of the CMB decreases. Place a clock at rest in every galaxy and set it to some agreed on time when some agreed on temperature 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 a galaxy so that it does not change in time: the particular r coordinate assigned at time t_0 is also assigned at other times. Isotropy ensures that all galaxies at a given t and r are at the same distance from the origin.

The coordinates (r, ϕ, θ) are said to be *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 the metric of a unit sphere is $d\,\Omega^2 = d\phi^2 + \sin^2\!\phi\, d\theta^2$. If the radius of the balloon at time t is $S(t)$, then the balloon's metric at time t is

$$ds^2 = S^2(t)\, d\,\Omega^2. \tag{4.2}$$

If $S(t)$ increases, then the balloon expands. Distances between the glued on galaxies increase, but their comoving $(\phi,\, \theta)$ 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

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

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

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

Robertson-Walker Metric

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

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

The metric was discovered independently by H. P. Robertson and E. W. Walker in the mid 1930's. You should compare it 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

$$S(t)\,d\sigma. \tag{4.5}$$

This is the distance measured by an inertial rigid rod. The distance is the product of two factors: $d\sigma$, which is independent of 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 do all distances between galaxies.

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

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

and contains the volume

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

If $k = -1$ (negative spatial curvature), then r is restricted to $0 \le r < 2$. The substitution $\rho = 2\tanh\xi$ shows that $V(2) = \infty$, i.e., the universe has infinite volume. If $k = 0$ (zero spatial curvature, flat), $0 \le r < \infty$ 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 $r \to \infty$. 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 has finite area and 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 Expansion Effects

Hubble's Law. At time t the distance from Earth at $r = 0$ to a galaxy at $r = r_e$ is, by Eq. (4.5),

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

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. (Yes, I know this is not a practical way to measure d!) The distance d increases with S.

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 Hubble's constant is given by

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

We see that Hubble's "constant" $H(t)$ depends on t. But for a fixed t (e.g., today), $v = Hd$ holds for all distances and directions.

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.

The Expansion Redshift. We 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 a light pulse is 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

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

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 $\Delta t_e/S(t_e) = \Delta t_o/S(t_o)$. Use Eq. (1.6) to obtain the desired relation:

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

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 $z = \infty$.

The most distant galaxy ever seen is at redshift $z = 8.68$, only 570 million years after the big bang. Gravitational lensing brightened the galaxy, perhaps by about 10Distances between galaxies were $1/9.7$ – about 10% – of what they are today when the light we see from the galaxy was emitted (Eq. (4.11)). The observed frequency of light from the galaxy is 10% of the emitted frequency (Eq. (1.7)). The observed rate of any other physical process detected at that distance, would also be 10% of the actual rate (Eq. (1.6)).

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

CMB Temperature. The CMB today is blackbody radiation at temperature $T(0) = 2.7\,\mathrm{K}$. According to Planck's law, its intensity at frequency f_o is proportional to

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

where h is Planck's constant and k is Boltzmann's constant. At distance z the radiation was at frequency f_e, where by Eq. (1.7), $f_e/f_o = z+1$. Thus the intensity of the radiation at z was proportional to

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

where

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

The spectrum was also blackbody, at temperature $T(z)$.

As stated in Sec. 4.1, the CMB was emitted at $\approx 3000\,\mathrm{K}$. Thus from Eq. (4.12), its redshift $z \approx 3000\,\mathrm{K}/2.7\,\mathrm{K} \approx 1100$.

Analysis of spectral (absorption) lines from a molecular hydrogen cloud at $z = 3$ backlit by a more distant quasar shows that the temperature of the CMB at $z = 3$ was $\sim 11\,\mathrm{K}$ CMB. Eq. (4.12) predicts this: $T(3) = (3+1)T(0) \approx 11\,\mathrm{K}$.

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 observational evidence that $k = 0$, i.e., that our universe is spatially flat. Together with other data, this will force us to modify the field equation. We then solve the modified field equation to determine $S(t)$.

The Universe Is Flat. The CMB provides evidence that our universe is spatially flat. When the CMB 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 (a few parts in 10^5) in the CMB. Figure 4.5, from the Planck spacecraft, shows colder regions in blue, and hotter in red.

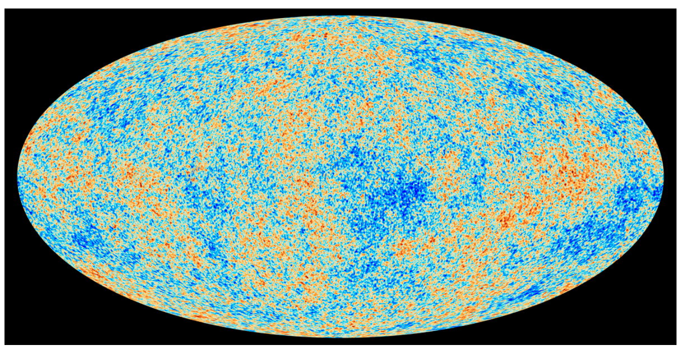


Fig. 4.5: Temperature fluctuations in the CMB.

The size of these regions does not depend on the curvature of space, but their angular size observed from Earth does. In a flat universe, the average temperature difference as a function of angular separation peaks at about .8 arcdegree. Observations in 2000 found the peak at .8 arcdegree, providing compelling evidence for a spatially (nearly) 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. 4.6 and suppose that light travels along the great circles of the sphere. According to Exercise 2.10 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 crosses 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 *angular size-redshift* relation gives the relationship between the angular size and redshift of objects in an expanding universe. It is derived in Appendix 16 and graphed in Fig. 4.9. We shall see that the middle graph in the figure corresponds to our universe. It has a minimum at $z = 1.6$.

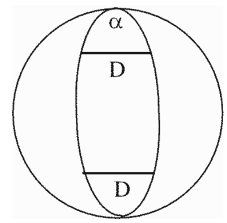


Fig. 4.6: Angular size on a sphere.

The Field Equation. Despite the strong evidence for a flat universe, we shall see that the field equation $\mathbf{G} = -8\pi\kappa\mathbf{T}$, Eq. (2.28), 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:

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

where Λ is a constant, called the *cosmological constant*. This is the most natural change possible to the field equation. In fact, the term $\Lambda\mathbf{g}$ appears in the field equation Eq. (A.16). It is eliminated there by assuming that spacetime is flat in the absence of matter. Thus the new term represents a gravitational effect of empty space; it literally "comes with the territory".

The cosmological constant Λ is so small that its effects in the solar system cannot be measured with today's technology. But we shall see that it dramatically affects the expansion of the universe.

To apply the new field equation, we need the energy-momentum tensor $\mathbf{T}$. In Sec. 4.2 we spread the matter in the universe into a uniform distribution of dust. Thus $\mathbf{T}$ is of the form Eq. (2.27). For the comoving matter, $dr = d\theta = d\phi = 0$, and from Eq. (4.4), $dt/ds = 1$. Thus $T^{tt} = \rho(t)$ 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:

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

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

Before solving the equations for S, we explore some of their consequences as they stand.

From Eqs. (4.14) and (4.9),

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

where

$$\rho_c = \frac{3H^2}{8\pi\kappa} \tag{4.17}$$

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. This is 10^{-7} as dense than the best vacuum created on Earth.

The term $\Lambda/8\pi\kappa$ in Eq. (4.16), like the other two terms, has the dimensions of a density. It is the density of a component of our universe called *dark energy*. We'll say more about dark energy soon. According to Eq. (4.16) the density of matter and dark energy add to ρ_c. But general relativity does not determine their individual values. They have been determined by observation. We discuss the results, first for matter and then for dark energy.

Matter. Two accurate determinations of the matter density ρ were announced in 2003. One used precise measurements of the anisotropies in the CMB. The other used measurements of the 3D distribution of hundreds of thousands of galaxies. Recent refinements of these two very different kinds of measurements give $\rho = .32\rho_c$. Some of this matter is in known forms such as protons, neutrons, electrons, neutrinos, and photons. But most is in a mysterious component of our universe called *dark matter*.

The measurements show that known forms of matter contribute only $.05\rho_c$ to ρ. This is consistent with two earlier, less accurate, determinations. One used spectra of distant galactic nuclei to "directly" measure the density of known forms of matter in the early universe. The other used the theory of big bang nucleosynthesis, which expresses the density of ordinary matter as a function of the abundance of deuterium in the unverse. This abundance was measured in distant pristine gas clouds backlit by even more distant quasars.

It is gratifying to have excellent agreement between the CMB and deuterium measurements, as the deuterium was created a minute or so after the big bang and the CMB was emitted hundreds of thousands of years later. Today over 90% of this matter is intergalactic hydrogen. Only a small fraction is in stars.

Dark matter contributes the remaining $.27\rho_c$ to ρ. Astronomers agree that dark matter exists and is five times as abundant as ordinary matter. Speculations abound, but no one knows what it is. Since ordinary and dark matter enter the field equations in the same way, via ρ in Eq. (4.14), they behave the same gravitationally.

Dark matter is required to prevent galaxies from flying apart as they rotate; there is insufficient ordinary matter to hold them together gravitationally. In fact, galaxies would not have formed without dark matter. They formed from the small density variations seen in the CMB. Any density variations of ordinary matter in the early universe were smoothed out by interactions with intense electromagnetic radiation. Dark matter neither emits nor absorbs electromagnetic radiation (hence its name). Thus variations within it could survive and grow by gravitational attraction. Computer simulations of the formation and evolution of galaxies from the time of the CMB to today produce the distribution of galaxies we see today – but only if they incorporate dark matter.

The best direct evidence for the existence of dark matter comes from spectacular observations on a vast scale: collisions of two galaxy clusters. About 100 million years ago, two clusters about 3.4 billion light years away collided and passed through each other. The result is the *bullet cluster*, shown in Fig. 4.7.

Fig. 4.7: The bullet cluster.

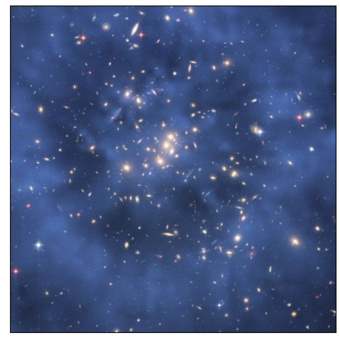


Fig. 4.8: A dark matter ring.

Intergalactic hydrogen gas and dark matter contribute most of the mass of the clusters. (The galaxies themselves contribute only a small amount.) During the collision, the gas in each cluster was slowed and heated by a drag force, similar to air resistance. Dark matter should slow less than the gas, since interacts with itself and ordinary matter only gravitationally. Thus the gas should lag behind the dark matter.

Figure 4.7 shows this. The very hot gas emits X-rays, shown in (false color) red. Dark matter is shown in blue. The distribution of this unseen matter was inferred by gravitational lensing of background objects.

A much older (1–2 billion years old) and farther away (5 billion light-years away) collision of two galaxy clusters provides similar evidence for dark matter. Unlike the bullet cluster collision, which we see from the side, this collision was along our line of sight. Dark matter "splashed" away from the collision to form the ring seen in Fig. 4.8. Again, the distribution of the unseen dark matter was inferred by gravitational lensing of background objects.

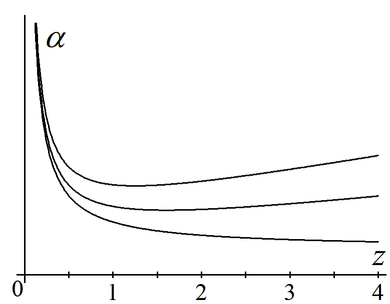


Fig. 4.9: The angular size-redshift relation. Upper: $\Lambda = 0$; middle: $\Lambda/8\pi\kappa = .68\rho_c$; lower: $\Lambda/8\pi\kappa = \rho_c$.

Dark Energy. The first evidence for the existence of dark energy came from observations of distant suprenovae made in the late 1990's. 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 depends on Λ. Fig. 4.10 graphs $\log(\ell)$ vs. z for a fixed L and three values of Λ.

Fig. 4.11 plots luminosity-redshift data for some supernovae with $0 \leq z \leq 1$. (The vertical line with each data point is an error bar.) The scale of the vertical axis is chosen so the $\Lambda/8\pi\kappa = .68\rho_c$ curve in Fig. 4.10 becomes the horizontal line in Fig. 4.11. The $\Lambda = 0$ curve in Fig. 4.10 becomes the dotted curve in Fig. 4.11. The data gives strong evidence for the existence of dark energy.

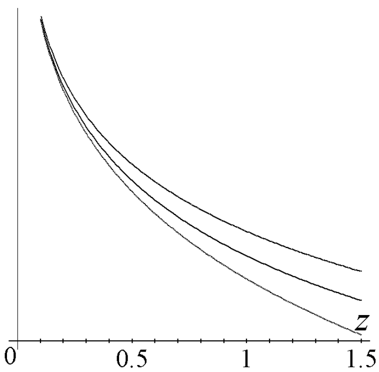


Fig. 4.10: The luminosity-redshift relation. Upper: $\Lambda = 0$; middle: $\Lambda/8\pi\kappa = .68\rho_c$; lower: $\Lambda/8\pi\kappa = \rho_c$.

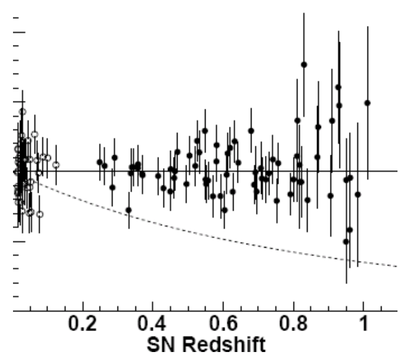


Fig. 4.11: Supernova luminosity-redshift data.

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

Independent evidence for dark energy comes from studies of galaxy clusters. For most of the history of the universe, clusters have grown in size by gravitational attraction. But with the increasing importance of the repulsive nature of dark energy in recent times (see below), the growth has slowed. Measurements of the growth of galaxy clusters out to $z = .9$ clearly show the effect of dark energy, with a value consistent with that given above.

The effect of dark energy has been probed to greater distances using certain very compact sources of radio frequency radiation. The *angular size-redshift relation*, also derived in Appendix 16, expresses the angular size α of an object as a function of its redshift. It is graphed in Fig. 4.9. These sources lie close to the curve for $\Lambda/8\pi\kappa = .68\rho_c$, out to $z = 3$.

Independent lines of evidence for the existence and density of dark energy come from the CMB, galaxy surveys, and gravitational lensing.

Matter + Dark Energy. The densities of matter, $\rho = .25\rho_c$, and dark energy, $\Lambda/8\pi\kappa = .68\rho_c$, add to ρ_c, as required by Eq. (4.16). Thus they are in accord with our new field equation, Eq. (4.13).

The discovery of dark energy caused a sensation. It was a turning point for cosmology. Science Magazine, one of the two premier science magazines in the world, called it the "Breakthrough of the Year" for 1998. Before the discovery, the standard belief was that the universe is flat and that $\Lambda = 0$. Equation (4.16) then requires that $\rho = \rho_c$. But enough matter to make this so was nowhere to be found. With the discovery of dark energy, suddenly everything fit. The standard cosmological model described in this chapter, whose parameters are known within a percent or two, is the result.

The Expansion Accelerates. A quick calculation using Eq. (4.15) shows that $\left(SS'^{\,2} - \Lambda S^{\,3}/3\right)' = 0$. Thus from Eq. (4.14),

$$\rho\, S^{\,3} = \text{constant}. \tag{4.18}$$

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. If we think of the surface of our balloon as representing dark energy, then it doesn't get thinner as it expands.

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.

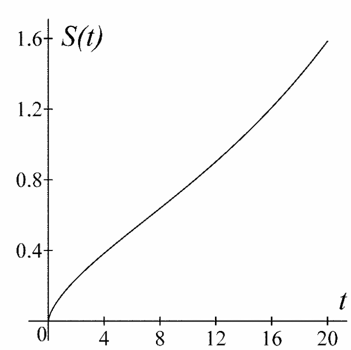


Fig. 4.12: S (in units of $S(t_o)$) as a function of t (in 10^9 years).

According to the luminosity-redshift relation, supernovae at $z \approx .5$ are $\sim$30% dimmer than they would be if Λ were 0. They are dimmer because their light has had to travel farther to get to us, as we accelerate away from them. Supernovae from $z = 1$ to $z \approx 1.5$ (the limit of the data) show the opposite effect: they are brighter than they would be had the expansion always been accelerating.

The changing rate of the expansion of the universe over its history is usually expressed as $H(z)$. Major efforts are underway to measure $H(z)$ at various distances to see if it conforms to the predictions of general relativity with a cosmological constant. Special attention is being given to determining the *transition time* from a decelerating to an accelerating expansion.

The Expansion Factor S(t). We now solve the field equations (4.14) and (4.15) for $S(t)$. Multiply Eq. (4.14) by $-S^3$ and use Eq. (4.18):

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

where a "o" subscript means that the quantity is evaluated today. Solve:

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

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. $S(t)$ is graphed in Fig. 4.12. As expected, the universe decelerates at early times and accelerates at later times.

Exercise 4.5. Show that Eq. (4.20) is a solution to Eq. (4.19). Hint: Use the square root of Eq. (4.20).

Exercise 4.6. Show that Eq. (4.20) implies that the age of the universe is $t_o = 13.4$ billion years.

Exercise 4.7. Show that the universe will expand exponentially as $t \to \infty$.

Exercise 4.8. Show that the decelerating/accelerating transition of the universe occurred at $z \approx .7$. That is, the light we receive today from galaxies at $z = .8$ was emitted at the time of the transition. That was about 7 billion years ago. Observations have confirmed this within about 10%.

Exercise 4.9. Find the expression from which Fig. 4.4 is graphed.

Summary. Today's *standard cosmological model* is based on general relativity with a cosmological constant: The universe had a big bang origin 13.8 billion years ago. (The Earth is 1/3 the age of the universe – 4.6 billion years old.) The universe is spatially flat, expanding, and the expansion is accelerating. It consists of 5% known matter, 27% dark matter, and 68% dark energy. Each of these numbers is supported by independent lines of evidence.

The construction of a model which fits the observations so well is a magnificent achievement. But it crystallized only at the turn of the millennium, with the discovery of the accelerating expansion. And we are left with two mysteries: dark matter and dark energy. They constitute 96% of the "stuff" of the universe, and we do not know what they are. 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.

Horizons. Light emitted at the big bang has only traveled 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 CMB.)

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 4.10. Show that with $S(t)$ from Eq. (4.20) this distance is ≈ 45 billion light years.

Actually, the distance to the horizon is somewhat uncertain, as it depends on $S(t)$ in the first moments after the big bang, for which Eq. (4.20) is not accurate, and which is uncertain.

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 $t_1 \to \infty$. 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 $t \to \infty$. Our *event horizon* is the sphere from which light emitted today reaches us as $t \to \infty$. We will never see galaxies beyond the event horizon as they are today because light emitted by them today will never reach us.

Exercise 4.11. 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 in this section. 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.

	$S'' = 0$	$v = c$	Min $\angle$ size	Event horizon	CMB	Particle horizon	Farthest ever seen
z	.78	1.5	1.6	1.8	1100	∞	–
d	8.9	13.6	14.5	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 discussion of Robertson–Walker spacetimes at the end of Sec. 4.2 illustrates 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 CMB, neutron stars, the double pulsar, gravitational lenses, the emission of gravitational radiation, black holes, detailed structure in the CMB, and the accelerating expansion of the universe. 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: It reduces to special relativity, an abundantly verified theory, in the absence of significant gravitational fields. It reduces to Newton's theory of gravity, an exceedingly accurate theory, in weak gravity with small velocities. 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 emission of gravitational radiation from the double pulsar system provide quantitative confirmation of the *full* field equation. And the application of general relativity on the grandest possible scale, to the universe as a whole, appears to be successful.

All this is very satisfying. But more tests of a fundamental theory are always welcome. The difficulty in finding tests for general relativity is that gravity is weak (about 10^{-40} as strong as electromagnetic forces), 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 soon complement electromagnetic waves in providing information about the universe. Sec. 3.5 described observations of the double pulsar which show indirectly that gravitational waves exist. The planning, construction, and commissioning of several gravitational wave detectors is under way. The waves are extremely weak and so are difficult to detect. One detector is able to measure changes in the distance between two test masses of 1/1000 of the diameter of a proton! We await the first 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). The theory of big bang nucleosynthesis uses the standard model. However, the standard model leaves many matters unexplained in cosmology. For example, dark matter and dark energy are beyond the scope of the theory. We may be on the verge of detecting dark matter directly in accelerator and other experiments.

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 – if they exist.

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. Success is by no means assured.

The general theory of relativity is a necessary and formidable 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.

Chapter A

Appendix

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

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

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

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

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

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

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

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

Eccentricity Mercury orbit = .206

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

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

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

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

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

$H_o = 2.3 \times 10^{-18}$/sec = (21 km/sec)/$10^6$ 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. Clock Synchronization. We show that clocks in an inertial frame can be synchronized according to Einstein's definition, Eq. (1.2). We will need an auxiliary assumption.

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. According to Exercise 1.5, $2T$ is independent of the time the light is emitted. 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. A.1. 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 4.

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'_Q, t'_P, t'_R. See Fig. A.1. We have the algebraic identities

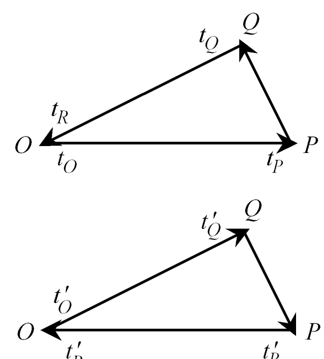


Fig. A.1: Light traversing a triangle in opposite directions.

$$
\begin{aligned}
t_R - t_O &= (t_R - t_P) + (t_P - t_Q) + (t_Q - t_O) \\
t'_R - t'_O &= (t'_R - t'_P) + (t'_P - t'_Q) + (t'_Q - t'_O).
\end{aligned}
\tag{A.1}
$$

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,

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

Thus, subtracting the equations Eq. (A.1) shows that the clocks at P and Q are synchronized:

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

A.4. 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.2.

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}.

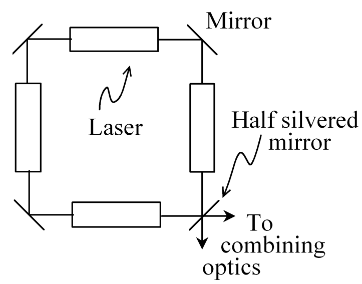


Fig. A.2: Comparing the round trip speed of light in opposite directions.

(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.)

A.5. Spacelike Separated Events. We show that if $c = 1$ in all inertial frames, then the proper distance formula Eq. (1.12) holds.

Let E and F be spacelike separated events with coordinate differences $(\Delta t, \Delta x)$ in an inertial frame I. For convenience, let E have coordinates $E(0,0)$. Then F has coordinates $F(\Delta t, \Delta x)$. By definition, the 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.

Exercise A.1. Show that E and F are simultaneous in an inertial frame in which the rod is at rest.

Since the events are simultaneous in an inertial frame, they are neither lightlike separated (so $|\Delta x| \neq |\Delta t|$), nor timelike separated (so $|\Delta x| \not< |\Delta t|$). Thus $|\Delta x| > |\Delta t|$, i.e., $|\Delta t/\Delta x| < 1$.

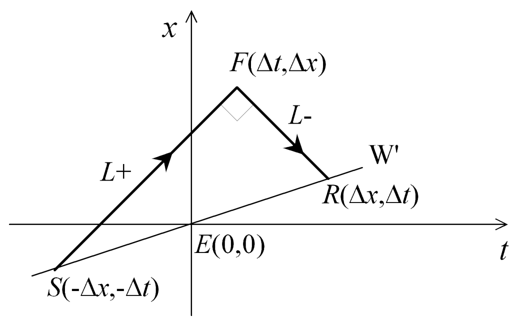


Fig. A.3: $|\Delta s| = (\Delta x^2 - \Delta t^2)^{\frac{1}{2}}$ for spacelike separated events E and F.

Fig. A.3 shows the worldline W' of an inertial observer O' moving with velocity $v = \Delta t/\Delta x$ (that is not a typo) in I. $L\pm$ 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 $R(\Delta x, \Delta t)$. (The time coordinate is Δx; the spatial coordinate is Δt.) Similarly, the equations for $L+$ and W' give the coordinates $S(-\Delta x, -\Delta t)$.

According to the time dilation formula Eq. (1.10), the proper time, as measured by O', between S and E and between E and R is $(\Delta x^2 - \Delta t^2)^{\frac{1}{2}}$. Since the times are equal, E and F are 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 $(\Delta x^2 - \Delta t^2)^{\frac{1}{2}}$. 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. Eq. (1.12) follows:

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

A.6. 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 symmetry.

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'.

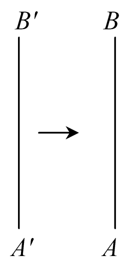


Fig. A.4: Rod perpendicular to its direction of motion in I.

Perpendicular Case. First, situate the rods as in Fig. A.4. 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. Symmetry demands that R and R' have the same length. For if, say, B' passed below B, then we would not have a symmetric situation: a rod moving in I is shorter, while an identical rod moving in I' is longer.

Parallel Case. Now situate the rods as in Fig. A.5. 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 symmetry: in each frame the moving rod is shorter.

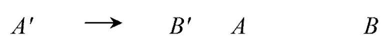


Fig. A.5: Rod parallel to its direction of motion in I.

Nor is there a 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.4, where the lengths are compared directly without clocks. Then there can be no disagreement over the relative lengths of the rods.

We now calculate the length contraction of R' in I. Let L' be the rest length of R' in I' and L be its length in I. A measurement of L' requires only a ruler, but a measurement of L also requires the synchronized clocks of I. For L is the distance between the positions of A' and B' in I, the positions being taken *at the same time* in I.

Emit a pulse of light from A' toward B'. Let v be the speed of R' in I. 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 $L/(1 - v)$ to reach B'. Similarly, if the pulse is reflected at B' back to A', it will take time $L/(1 + v)$ to reach A'. Thus the round trip time in I is

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

According to Eq. (1.11), a clock at A' will measure a total time

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

for the pulse to leave and return.

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

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

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

The factor $(1 - v^2)^{\frac{1}{2}}$ appears in both the proper distance formula Eq. (1.12) and the length contraction formula above. In both cases, a rod moving in I is at rest in I'. In both cases, events at which the ends of the rod are present simultaneously are considered. Nevertheless, the cases are different: for proper distance the events are simultaneous in I', whereas for length contraction the events are simultaneous in I.

A.7. 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.6.

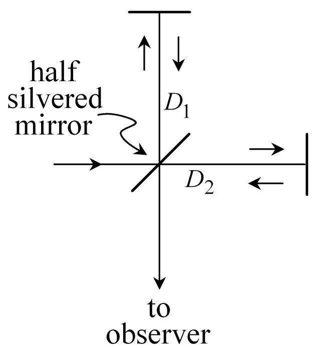


Fig. A.6: A Michelson interferometer.

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

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

cycles out of phase.

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

$$N - N' = 2f\ (D_1 + D_2)\left(\frac{1}{c_1} - \frac{1}{c_2}\right) \approx 2f\,\frac{D_1 + D_2}{c}\frac{\Delta c}{c}. \tag{A.3}$$

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

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

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.7 shows a schematic diagram of the experiment.

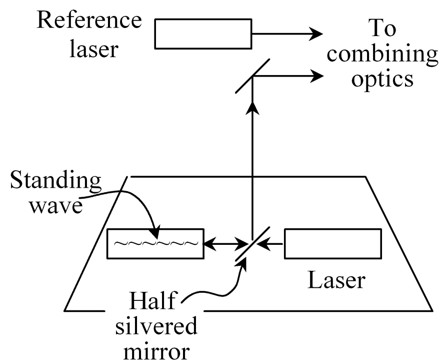


Fig. A.7: The Brillit-Hall experiment.

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

$$Tf = N. \tag{A.4}$$

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.4) to give $Df = cN$. 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}.

As of 2009 the best result is that the speed of light in different directions differs by less than 1 part in 10^{17}.

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.2), we obtain

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

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

A.8. Newtonian Orbits. Let the Sun be at the origin of a coordinate system. Let the vector $\mathbf{r}(t)$ describe the path of a planet. At every point $\mathbf{r}$ define orthogonal unit vectors

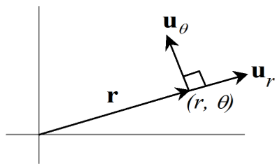


Fig. A.8: The polar unit vectors.

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

See Fig. A.8. 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

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

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

$$r^2\frac{d\theta}{dt} = A\,, \tag{A.5}$$

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.5), 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:

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

where the vector $\mathbf{e}$ is a constant of integration. Take the inner product of this with $\mathbf{u}_\theta$ and use Eq. (A.5) again:

$$\frac{\kappa M}{A^2}\left[1 + e\cos\left(\theta - \theta_p\right)\right] = \frac{1}{r}\,, \tag{A.6}$$

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 $\theta = \theta_p$.

A.9. 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 $\big(\bar{f}_{mn}(\bar{x})\big)$ 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 $g_{ji}\Gamma^i_{kp} = \frac{1}{2}\left(\partial_p g_{kj} + \partial_k g_{jp} - \partial_j g_{kp}\right)$. Exchange j and k to obtain $g_{ki}\Gamma^i_{jp} = \frac{1}{2}\left(\partial_p g_{jk} + \partial_j g_{kp} - \partial_k g_{jp}\right)$. Add the two equations to obtain the desired expression:

$$\partial_p g_{jk} = g_{ji}\Gamma^i_{kp} + g_{ki}\Gamma^i_{jp}\,. \tag{A.7}$$

Define new coordinates $\bar{x}$ by

$$x^m = \bar{x}^m - \tfrac{1}{2}\Gamma^m_{st}\bar{x}^s\bar{x}^t\,,$$

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

$$\frac{\partial x^m}{\partial \bar{x}^j} = \frac{\partial \bar{x}^m}{\partial \bar{x}^j} - \tfrac{1}{2}\Gamma^m_{sj}\bar{x}^s - \tfrac{1}{2}\Gamma^m_{jt}\bar{x}^t = \delta^m_j - \Gamma^m_{sj}\bar{x}^s\,.$$

Substitute this into Eq. (2.9):

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

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.7) to $\mathbf{f}$:

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

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

A.10. 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

$$y^i_r = \frac{\partial y^i}{\partial x^r}\,,\; x^u_v = \frac{\partial x^u}{\partial y^v}\,,\; y^i_{rs} = \frac{\partial^2 y^i}{\partial x^s \partial x^r}\,,\; x^u_{vw} = \frac{\partial^2 x^u}{\partial y^w \partial y^v}\,.$$

Differentiate $\dot y^i = y^i_q \dot x^q$ and use $\ddot x^q = 0$, Eq. (2.16):

$$\ddot y^i = y^i_q \ddot x^q + y^i_{qr} \dot x^r \dot x^q = y^i_{qr}(x^r_j \dot y^j)(x^q_k \dot y^k)\,.$$

In terms of components, the formula $\mathbf{g}^{-1} = \mathbf{a}^{-1}\,\mathbf{f}^{\circ-1}\left(\mathbf{a}^{-1}\right)^t$ from Exercise 2.6c is

$$g^{im} = y^i_r f^{\circ\,rs} y^m_s\,.$$

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:

$$\partial_i g_{jk} = (\partial_p f_{mn})\, x^p_i x^m_j x^n_k + f_{mn} x^m_{ji} x^n_k + f_{mn} x^m_j x^n_{ki} = 2 f^\circ_{mn} x^m_j x^n_{kj}\,.$$

Note that $(x^u_m y^m_s) = (\partial x^u / \partial x^s)$ is the identity matrix. Substitute the last two displayed equations into the definition Eq. (2.17) of the Christoffel symbols:

$$\begin{aligned}
\Gamma^i_{jk} &= \tfrac{1}{2}\, g^{im}\left(\partial_k g_{jm} + \partial_j g_{mk} - \partial_m g_{jk}\right)\\
&= y^i_r f^{\circ\,rs} y^m_s \left(f^\circ_{uv} x^u_j x^v_{mk} + f^\circ_{uv} x^u_m x^v_{kj} - f^\circ_{uv} x^u_j x^v_{km}\right)\\
&= y^i_r f^{\circ\,rs} y^m_s f^\circ_{uv} x^u_m x^v_{kj}\\
&= y^i_r f^{\circ\,rs} f^\circ_{uv} x^u_m y^m_s x^v_{kj}\\
&= y^i_r f^{\circ\,rs} f^\circ_{sv} x^v_{kj}\\
&= y^i_r x^r_{kj}\,.
\end{aligned}$$

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

$$\ddot y^i + \Gamma^i_{jk}\, \dot y^j \dot y^k = \left(y^i_{qr} x^q_k x^r_j + y^i_r x^r_{kj}\right) \dot y^j \dot y^k = \partial_k\left(y^i_r x^r_j\right) \dot y^j \dot y^k = 0\,.$$

A.11. Tensors. Tensors are a generalization of vectors. (See Exercise A.3.) 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} and the Ricci tensor R_{ij} have two. Another example is the *Riemann (curvature) tensor*, which has four:

$$R^p{}_{rsq} = \Gamma^p_{ts}\,\Gamma^t_{rq} - \Gamma^p_{tq}\,\Gamma^t_{rs} + \partial_s\Gamma^p_{rq} - \partial_q\Gamma^p_{rs}\,. \tag{A.8}$$

A superscript index is called *contravariant*; a subscript index is called *covariant*.

A *scalar* is a tensor with a single component, i.e., it is a number, and so requires no indices. A scalar has the same value in all coordinate systems.

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. Nevertheless, we give a short introduction here, for those who are curious.

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.

The components of a vector change under a rotation of coordinates. Similarly, the components of a tensor change under a change of coordinates. 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 tensors. However, it is the best we can do without getting into heavier mathematics.

Exercise A.2. Show that the coordinate differences (dy^i) form a tensor.

Exercise A.3. This exercise shows that vectors in $\mathbf{R}^n$ are tensors. Fix a point P and think of vectors $\mathbf{v}$ as extending from P as origin. Let (y^i) be a global coordinate system. Define a basis $(\boldsymbol{dy}^j)$ at P by $\boldsymbol{dy}^j = \nabla y^j|_P$. The vector $\boldsymbol{dy}^j$ is orthogonal to the surface $y^j =$ constant through P. Expand $\mathbf{v}$ with respect to this basis: $\mathbf{v} = v_j\,\boldsymbol{dy}^j$. Show that the v_j transform as a covariant tensor. Hint: Let (x^i) be a rectangular coordinate system with orthonormal basis $(\mathbf{e}_i)$. Then $\boldsymbol{dy}^j = (\partial y^j/\partial x^i)\,\mathbf{e}_i$. Equate the $\mathbf{e}_i$ coefficients of $v_j\,\boldsymbol{dy}^j = \bar{v}_j\,\boldsymbol{d\bar{y}}^j$. Multiply both sides by $\partial x^i/\partial\bar{y}^k$.

A superscript index p can be *lowered* to a subscript index t by multiplying by g_{tp} and summing. For example, $R_{trsa} = g_{tp}R^p{}_{rsa}$. Similarly, an index can be *raised* using the inverse matrix (g^{pt}) : $R^p{}_{rsa} = g^{pt}R_{trsa}$.

Exercise A.4. Show that lowering and raising a tensor index are inverse operations. Use the example $R^p{}_{rsa}$ above.

Exercise A.5. The matrix (g^{mn}) is the inverse of the matrix (g_{jk}). But now g^{mn} also denotes g_{jk} with its indices raised. Prove that the two meanings do not conflict.

Exercise A.6. **a.** Show that the indices on g^{mn} are contravariant indices. Hint: Write Eq. (2.12) as a matrix equation and invert it. C.f. Exercise 2.6.
b. Show that a raised covariant index is a contravariant index.

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:

$$R_{jk} = R^p{}_{jkp}.$$

Exercise A.7. **a.** Show that the product of two tensors is a tensor. Use the example $g_{tu}R^p{}_{rsa}$ above.
b. Show that the contraction of a tensor is a tensor. Use the example $g_{tp}R^p{}_{rsa}$ above.

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 $(\ddot{y}^i + \Gamma^i_{jk}\,\dot{y}^j\dot{y}^k)$, *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.12. Fermi Normal Coordinates. Recall from Appendix 9 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

$$\begin{aligned} f_{00} &= \quad 1 \quad - \quad R_{0l0m}\, x^l x^m + \dots \\ f_{0i} &= \quad 0 \quad - \tfrac{2}{3} R_{0lim}\, x^l x^m + \dots \\ f_{ij} &= -\delta_{ij} - \tfrac{1}{3} R_{iljm}\, x^l x^m + \dots\ , \end{aligned} \tag{A.9}$$

where none of $i,\ j,\ l,\ m$ is 0; the R_{iljm} are components of the Riemann tensor, Eq. (A.8), 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 need 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:

$$\begin{gathered} R_{iljm} = -R_{ilmj},\ R_{iljm} = -R_{lijm}, \\ R_{iljm} = R_{imjl},\ R_{iljm} + R_{imlj} + R_{ijml} = 0. \end{gathered}$$

The first two identities imply that $R_{ilmm} = 0$ and $R_{iijm} = 0$.)

A calculation of the Einstein tensor from the expansion Eq. (A.9) gives

$$G^{00} = R_{1212} + R_{2323} + R_{3131}\,. \tag{A.10}$$

Eq. (2.23) gives the curvature K of a surface when $g_{12} = 0$. The general formula is

$$K = \frac{R_{1212}}{\det(\mathbf{g})}. \tag{A.11}$$

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.9) the surface has metric

$$f_{ij} = \delta_{ij} + \tfrac{1}{3} R_{iljm} x^l x^m + \dots\ ,$$

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.11), the curvature of the surface at the origin is

$$K_{12} = \frac{-R_{1212}}{\det(\delta_{ij})} = -R_{1212}. \tag{A.12}$$

Putting together Eqs. (A.10) and (A.12),

$$G^{00} = -(K_{12} + K_{23} + K_{31})\,.$$

A.13. The Form of the Field Equation. We give a plausibility argument, based on reasonable assumptions, which leads from the schematic field equation Eq. (2.24) to Einstein's field equation Eq. (2.28).

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

Consider first the right side of Eq. (2.24), "quantity determined by mass/energy". 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. (1.11), this factor is dx^0/ds. Second, from Appendix 6, 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

$$T^{00} = \rho \, \frac{dx^0}{ds} \frac{dx^0}{ds}, \tag{A.13}$$

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. (A.13) is only one component of a whole:

$$T^{jk} = \rho \, \frac{dx^j}{ds} \frac{dx^k}{ds}\,. \tag{A.14}$$

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. (A.14). Transforming to global coordinates, the right side of Eq. (2.24) is the energy-momentum tensor $\mathbf{T}$. Thus the field equation is of the form

$$\mathbf{G} = -\mathbf{8}\pi\kappa\mathbf{T}, \tag{A.15}$$

where according to Eq. (2.24) the left side is a "quantity determined by metric". We denote it $\mathbf{G}$ in anticipation that it will turn out to be the Einstein tensor. And $-8\pi\kappa$ was inserted for later convenience.

We shall make four assumptions which will uniquely determine $\mathbf{G}$.

(i) From Eq. (2.24), $\mathbf{G}$ depends on $\mathbf{g}$. Assume that the components of $\mathbf{G}$, like the curvature K of a surface, depend only on the components of $\mathbf{g}$ 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^{th} 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 Eq. (A.15) is of the form

$$A\left[\mathbf{R} - \tfrac{1}{2}\,\mathbf{g}R\right] + \Lambda\mathbf{g} = -8\pi\kappa\mathbf{T}\,, \tag{A.16}$$

where A and Λ are constants.

It remains only to determine Λ and A.

(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. (A.16) becomes $\Lambda \mathbf{g} = 0$. Therefore $\Lambda = 0$.

(iv) Assume that Einstein's theory reduces to Newton's when the latter is accurate, namely when gravity is weak and velocities are small. We will show that this implies that $A = 1$. Then Eq. (A.16) will be precisely the field equation Eq. (2.28).

To show that $A = 1$, we use results from Chapter 4 with $\Lambda = 0$. Solve Eq. (4.15) for S'' and use Eq. (4.14):

$$S'' = -\frac{4\pi\kappa\rho S}{3}\,. \tag{A.17}$$

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 Schwartzschild 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.17); Einstein's and Newton's theories give the same result.

But if $A \neq 1$ in Eq. (A.16), then the left side of Eq. (A.17) would be multiplied by A. Thus the coincidence of the two theories in this situation requires $A = 1$.

A.14. Approximating an Integral. We approximate the integral, Eq. (3.26). Retaining only first order terms in the small quantities m/r and m/r_s we have

$$\begin{aligned}
\frac{(1-2m/r)^{-2}}{1-\dfrac{1-2m/r}{1-2m/r_s}(r_s/r)^2} &\approx \frac{(1+2m/r)^2}{1-(1-2m/r)(1+2m/r_s)(r_s/r)^2} \\
&\approx \frac{1+4m/r}{1-(1-2m/r+2m/r_s)(r_s/r)^2} \\
&= \frac{1+4m/r}{\left(1-(r_s/r)^2\right)\left(1-\dfrac{2mr_s}{r(r+r_s)}\right)} \\
&\approx \left(1-\left(\frac{r_s}{r}\right)^2\right)^{-1}\left(1+\frac{4m}{r}\right)\left(1+\frac{2mr_s}{r(r+r_s)}\right) \\
&\approx \left(1-\left(\frac{r_s}{r}\right)^2\right)^{-1}\left(1+\frac{4m}{r}+\frac{2mr_s}{r(r+r_s)}\right).
\end{aligned}$$

We now see that Eq. (3.26) can be replaced with

$$t = \int_{r_e}^{r_s}\left(1-\left(\frac{r_s}{r}\right)^2\right)^{-1}\left(1+\frac{4m}{r}+\frac{2mr_s}{r(r+r_s)}\right)dr\,.$$

Integration gives

$$t = \left(r_e^2-r_s^2\right)^{\frac{1}{2}} + 2m\ln\frac{r_e+\left(r_e^2-r_s^2\right)^{\frac{1}{2}}}{r_s} + m\left(\frac{r_e-r_s}{r_e+r_s}\right)^{\frac{1}{2}}\,.$$

Use $r_s \ll r_e$ to obtain Eq. (3.27).

A.15. The Robertson-Walker Metric. This appendix derives the Robertson-Walker metric for an isotropic universe, Eq. (4.4). The field equation is not used.

From Eq. (3.2) and Exercise 3.3, the metric of a spherically symmetric spacetime can be put in the form

$$ds^2 = \mathrm{e}^{2\mu} dt^2 - \mathrm{e}^{2\nu} dr^2 - r^2 \mathrm{e}^{2\lambda} d\,\Omega^2, \tag{A.18}$$

where the coefficients depend on R but not ϕ or θ. Since the universe expands, the coefficients can also depend on t, unlike those in Eq. (3.2).

Since the coordinates are comoving, $dr = d\phi = d\theta = 0$ for neighboring events on the worldline of a galaxy. Thus by Eq. (A.18), $ds^2 = \mathrm{e}^{2\mu} dt^2$. But $ds = dt$, since both are the time between the events measured by a clock in the galaxy. Thus $\mathrm{e}^{2\mu} = 1$. Our metric is now of the form

$$ds^2 = dt^2 - \left(\mathrm{e}^{2\nu} dr^2 - r^2 \mathrm{e}^{2\lambda} d\,\Omega^2\right) . \tag{A.19}$$

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 $S(t)$, as in the metric of our balloon, Eq. (4.2). We 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.19), $dt^2 = \mathrm{e}^{2\nu} dr^2$. Thus the light travel time between $P(r, \phi, \theta)$ and $Q(r + dr, \phi, \theta)$ is $\mathrm{e}^{\nu(t,r)} dr$. The ratio of this time to that measured at another time t' is given by $\mathrm{e}^{\nu(t,r)}/\mathrm{e}^{\nu(t',r)}$. Similarly, the ratio of the times between P and $R(r, \theta, \phi + d\phi)$ is $\mathrm{e}^{\lambda(t,r)}/\mathrm{e}^{\lambda(t',r)}$. By isotropy at P the ratios are equal:

$$\frac{\mathrm{e}^{\nu(t,r)}}{\mathrm{e}^{\nu(t',r)}} = \frac{\mathrm{e}^{\lambda(t,r)}}{\mathrm{e}^{\lambda(t',r)}} . \tag{A.20}$$

(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.20) 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.20) are a function only of t, say $S(t)$. Set $\nu(t', r) = \nu(r)$ to obtain $\mathrm{e}^{\nu(t,r)} = S(t)\,\mathrm{e}^{\nu(r)}$. Similarly, $\mathrm{e}^{\lambda(t,r)} = S(t)\,\mathrm{e}^{\lambda(r)}$. Thus the metric Eq. (A.19) becomes

$$ds^2 = dt^2 - S^{\,2}(t) \left(\mathrm{e}^{2\nu} dr^2 + r^2 \mathrm{e}^{2\lambda} d\,\Omega^2\right). \tag{A.21}$$

The coordinate change $\bar{r} = r\mathrm{e}^{\lambda(r)}$ puts Eq. (A.21) in the form

$$ds^2 = dt^2 - S^{\,2}(t) \left(\mathrm{e}^{2\nu} dr^2 + r^2 d\,\Omega^2\right) = dt^2 - S^{\,2}(t)\, d\sigma^2. \tag{A.22}$$

We now determine ν in Eq. (A.22).

Exercise A.8. 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 $\nu\,' \mathrm{e}^{-2\nu} = Kr$.

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.22) 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.9. Since the length of the subtended arc is the radius times the subtended angle, $\mathrm{e}^{\nu(0)} = 1$. Thus $C = 1$.

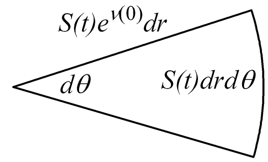


Fig. A.9: Evaluating the integration constant C.

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.

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

$$\left(\frac{S'}{S}\right)^2 = \frac{8\pi\kappa\rho(t)}{3} + \frac{\Lambda}{3} = H_o^2\left[\Omega_\Lambda + (1-\Omega_\Lambda)(z+1)^3\right], \tag{A.23}$$

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.23):

$$(1+z)^{-2}\left(\frac{dz}{dt}\right)^2 = H_o^2\left[\Omega_\Lambda + (1-\Omega_\Lambda)(z+1)^3\right]. \tag{A.24}$$

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.24), separate variables, and integrate to give the distance-redshift relation:

$$d(z) = S(t_o)r_e = H_o^{-1}\int_0^z \left[\Omega_\Lambda + (1-\Omega_\Lambda)(\zeta+1)^3\right]^{-\frac{1}{2}} d\zeta\,. \tag{A.25}$$

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.25), to obtain the angular size-redshift relation:

$$\alpha(z) = \frac{(z+1)H_o D}{\int\limits_0^z \left[\Omega_\Lambda + (1-\Omega_\Lambda)(\zeta+1)^3\right]^{-\frac{1}{2}} d\zeta} \,. \tag{A.26}$$

This is graphed in Fig. 4.9.

The Luminosity-Redshift Relation. The luninosity-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:

$$\ell = \frac{L}{4\pi(1+z)^2 d^{\,2}} \,. \tag{A.27}$$

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.25), into Eq. (A.27) to obtain the luninosity-redshift relation:

$$\ell(z) = \frac{H_o^2 L}{4\pi\,(1+z)^2 \left(\int\limits_0^z \left\{\Omega_\Lambda + (1-\Omega_\Lambda)(\zeta+1)^3\right\}^{-\frac{1}{2}} d\zeta\right)^2} \,. \tag{A.28}$$

This is graphed in Fig. 4.10.

Index