GENERAL RELATIVITY

A GEOMETRIC APPROACH

Malcolm Ludvigsen

University of Linköping

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http: //www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http: //www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1999

Printed in the United States of America

Typeset in Melior 9.75/13pt. and Eurostile in LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Ludvigsen, Malcolm. 1946–
General relativity : a geometric approach / Malcolm Ludvigsen.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-63019-3 (hardbound)
1. General relativity (Physics) 2. Space and time.
3. Geometrodynamics. I. Title.
QC173.6.L83 1999
530.11 – dc21 98-37546
CIP

ISBN 0 521 63019 3 hardback
ISBN 0 521 63976 x paperback

Contents

1

Introduction

One of the greatest intellectual achievements of the twentieth century is surely the realization that space and time should be considered as a single whole – a four-dimensional manifold called spacetime – rather than two separate, independent entities. This resolved at one stroke the apparent incompatibility between the physical equivalence of inertial observers and the constancy of the speed of light, and brought within its wake a whole new way of looking at the physical world where time and space are no longer absolute – a fixed, god-given background for all physical processes – but are themselves physical constructs whose properties and geometry are dependent on the state of the universe. I am, of course, referring to the special and general theory of relativity.

This book is about this revolutionary idea and, in particular, the impact that it has had on our view of the universe as a whole. From the very beginning the emphasis will be on **spacetime** as a single, undifferentiated four-dimensional manifold, and its physical geometry. But what do we actually mean by spacetime and what do we mean by its physical geometry?

A point of spacetime represents an **event**: an instantaneous, pointlike occurrence, for example lightning striking a tree. This should be contrasted with the notion of a **point in space**, which essentially represents the position of a pointlike particle with respect to some frame of reference. In the spacetime picture, a particle is represented by a curve, its **world line**, which represents the sequence of events that it "occupies" during its lifetime. The life span of a person is, for example, a sequence of events, starting with birth, ending with death, and punctuated by many happy and sad events. A short meeting between two friends is, for example, represented in the spacetime picture as an intersection of their world lines. The importance of the spacetime picture is that it does not depend on the initial imposition of any notion of **absolute time** or **absolute space** – an event is something in its own right, and we don't need to represent it by (t, p), where p is its absolute position and t is its absolute time.

The geometry of spacetime is not something given *a priori*, but something to be discovered from physical observations and general physical principles. A geometrical statement about spacetime is really a statement about physics, or a relationship between two or more events that any observer would agree exists. For example, the statement that two events A and B can be connected by a light ray is geometrical in that if one observer finds it to be true then all observers will find it to be true. On the other hand, the statement that A and B occur in the same place, like lightning striking the same tree twice, is certainly nongeometrical. Certainly the tree will appear to be in the same place at each lightning strike to a person standing nearby, but not to a passing astronaut who happens to be flying by in his spaceship at 10,000 miles per hour.

Just as points and curves are the basic elements of Euclidean geometry, events and world lines are the basic elements of spacetime geometry. However, whereas the rules of Euclidean geometry are given axiomatically, the rules of spacetime geometry are statements about the physical world and may be viewed as a way of expressing certain fundamental laws of physics. Just as in Euclidean geometry where we have a special set of curves called straight lines, in spacetime geometry we have a special set of world lines corresponding to freely moving (inertial) massive particles (e.g. electrons, protons, cricket balls, etc.) and an even more special set corresponding to freely moving massless particles (e.g. photons). In the next few chapters we shall show how the geometry of spacetime can be constructed from these basic elements. Our main guide in this endeavor will be the **principle of relativity**, which, roughly speaking, states the following:

If any two inertial observers perform the same experiment covering a small region of spacetime then, all other things being equal, they will come up with the same results.

In other words, all inertial observers are equal as far as the fundamental laws of physics are concerned. For example, the results of an experiment performed by an astronaut in a freely moving, perfectly insulated spaceship would give no indication of his spacetime position in the universe, his state of motion, or his orientation. Of course, if he opened his curtains and looked out of the window, his view would be different from that of some other inertial observer in a different part of the universe. Thus, not all inertial observers are equal with respect to their *environment*, and, as we shall see, the environment of the universe as a whole selects out a very special set of inertial observers who are, in a sense, in a state of absolute rest with respect to the large-scale structure of the universe. This does not, however, contradict the principle of relativity, because this state of motion is determined by the state of the universe rather than the fundamental laws of physics.

The most important geometrical structure we shall consider is called the **spacetime metric**. This is a tensorial object that essentially determines the distance (or time) between two nearby events. It should be emphasized that the existence of a spacetime metric and its properties are derived from the principle of relativity and the behavior of light; it is therefore a physical object that encodes certain very fundamental laws of nature. The metric determines another tensorial object called the **curvature tensor**. At any given event this essentially encodes all information about the gravitational field in the neighborhood of the event. It may also, in a very real sense, be interpreted as describing the curvature of spacetime. Another tensorial object we shall consider is called the **energy–momentum tensor**. This describes the mass (energy) content of a small region of spacetime. That this tensor and the curvature tensor are related via the famous **Einstein equation**

$$G_{ab} = -8\pi T_{ab}$$

is one of the foundations of general relativity. This equation gives a relationship between the curvature of spacetime (G_{ab}) and its mass content T_{ab}. Needless to say, it has profound implications as far as the geometry of the universe is concerned.

When dealing with spacetime we are really dealing with the very bedrock of physics. All physical processes take place within a spacetime setting and, indeed, determine the very structure of spacetime itself. Unlike other branches of physics, which are more selective in their subject matter, the study of spacetime has an all-pervading character, and this leads, necessarily, to a global picture of the universe as a whole. Given that the fundamental laws of physics are the same in all regions of the universe, we are led to a global spacetime description consisting of a four-dimensional manifold, M, whose elements represent all events in the universe, together with a metric g_{ab}. The matter content of the universe determines an energy-momentum tensor, T_{ab}, and this in turn determines the curvature of spacetime via Einstein's equation.

In the same way as the curvature of the earth may be neglected as long as we stay within a sufficiently small region on the earth's surface, the curvature of spacetime (and hence gravity) may be neglected as long as we restrict attention to a sufficiently small region of spacetime. This leads to a flat-space description of nature, which is adequate for situations where gravitational effects may be neglected. The study of flat spacetime and physical processes within such a setting is called **special relativity**. This will be described from a geometric point of view in the first few chapters of this book.

To bring gravity into the picture we must include the curvature of spacetime. This leads to a very elegant and highly successful theory of gravity known as **general relativity**, which is the main topic of this book. Not only is it compatible with Newtonian theory under usual conditions (e.g. in the solar system), but it yields new effects under more extreme conditions, all of which have been experimentally verified. Perhaps the most exciting thing about general relativity is that it predicts the existence of very exotic objects known as **black holes**. A black hole is essentially an object whose gravitational field is so strong as to prevent even light from escaping. Though the observational evidence is not yet entirely conclusive, it is generally believed that such objects do indeed exist and may be quite common.

In the final part of the book we deal with cosmology, which is the study of the large-scale structure and behavior of the universe as a whole. At first sight this may seem to be a rash and presumptuous exercise with little chance of any real success, and better left to philosophers and theologians. After all, the universe as a whole is a very complicated system with apparently little order or regularity. It is true that there exist fundamental laws of nature that considerably reduce the randomness of things (e.g., restricting the orbits of planets to be ellipses rather than some arbitrary curves), but they have no bearing on the *initial conditions* of a physical system, which can be – and, in real life, are – pretty random. For example, there seems to be no reason why the planets of our solar system have their particular masses or particular distances from the sun: Newton's law of gravity would be consistent with a very different solar system. Things are, however, much less random on a very small scale. The laws of quantum mechanics, for example, determine the energy levels of a hydrogen atom independent of any initial conditions, and, on an even smaller scale, there is hope that masses of all fundamental particles will eventually be determined from some very basic law of nature. The world is thus very regular on a small scale, but as we increase the scale of things, irregularity and randomness seem to increase.

This is true up to a point, but, as we increase our length scale, regularity and order slowly begin to reappear. We certainly know more about the mechanics of the solar system than about the mechanics of human interaction, and the structure and evolution of stars is much better understood than that of bacteria, say. As we increase our length scale still further to a sufficiently large galactic level, a remarkable degree of order and regularity becomes apparent: the distribution of galaxies appears to be spatially homogeneous and isotropic. Clearly, this does not apply to *all* observers – even inertial observers. If it were true for one observer, then, because of the Doppler effect, it would not be true for another observer with a high relative speed. We shall return to this point shortly.

It thus appears that the universe is not simply a random collection of irregularly distributed matter, but is a single entity, all parts of which are in some sense in unison with all other parts. This is, at any rate, the view taken by the **standard model** of cosmology, which will be our main concern.

The universe, as we have seen, appears to be homogeneous and isotropic on a sufficiently large scale. These properties lead us to make an assumption about the model universe we shall be studying, called the **cosmological principle**. According to this principle the universe is homogeneous everywhere and isotropic about every point in it. This assumption is very important, and it is remarkable that the universe seems to obey it. The universe is thus not a random collection of galaxies, but a single unified entity. As we stated above, the cosmological principle is not true for all observers, but only for those who are, in a sense, at rest with respect with the universe as a whole. We shall refer to such observers as being **comoving**. With this in mind, the cosmological principle may be stated in a spacetime context as follows:

- Any event E can be occupied by just one comoving observer, and to this observer the universe appears isotropic. The set of all comoving world lines thus forms a congruence of curves in the spacetime of the universe as a whole, in the sense that any given event lies on just one comoving world line.
- Given an event E on some comoving world line, there exists a unique corresponding event E' on any other comoving world line such that the physical conditions at E and E' are identical. We say that E and E' lie in the same ***epoch*** and have the same universal time t.

By combining the cosmological principle in this form with Einstein's equation we obtain a mathematical model of the universe as a whole, called the standard big-bang model, which makes the following remarkable predictions:

(i) The universe cannot be static, but must either be expanding or contracting at any given epoch. This is, of course, consistent with Hubble's observations, which indicate the universe is expanding in the present epoch.

(ii) Given that the universe is now expanding, the matter density $\rho(t)$ of the universe at any universal time t in the past must have been a decreasing function of t, and furthermore there exists a *finite* number t_0 such that

$$\lim_{t \to t_0^+} \rho(t) = \infty.$$

The density of the universe thus increases as we move back in time, and can achieve an arbitrarily large value within a *finite* time. From now on we shall choose an origin for t such that $t_0 = 0$.

(iii) The $t =$ constant cross sections corresponding to different epochs are spaces of constant curvature. If their curvature is positive (a closed universe) then the universe will eventually start contracting. If, on the other hand, their curvature is negative or zero (an open universe), then the universe will continue to expand forever.

(iv) Assuming that all matter in the universe was once in thermal equilibrium, then the temperature $T(t)$ would have been a decreasing function of t and

$$\lim_{t \to 0^+} T(t) = \infty.$$

In other words, the early universe would have been a very hot place.

(v) There will have existed a time in the past when radiation ceased to be in thermal equilibrium with ordinary matter. Though not in thermal equilibrium after this time, the radiation will have retained its characteristic blackbody spectrum and should now be detectable at a much lower temperature of about 3 K. Such a cosmic background radiation was discovered by Penzias and Wilson in 1965, thus giving a very convincing confirmation of the standard model. Furthermore, this radiation was found to be extremely isotropic, thus lending support to the cosmological principle.

(vi) Using the well-tried methods of standard particle physics and statistical mechanics, the standard model predicts the present abundance of the lighter elements in the universe. This prediction has been confirmed by observation. For a popular account of this see, for example, Weinberg (1993).

A very disturbing feature of the standard model is that it predicts that the universe started with a **big bang** at a *finite* time in the past. What happened before the big bang, and what was the nature of the event corresponding to the big bang itself? Such questions are based on deeply ingrained, but false, assumptions about the nature of time. The spacetime manifold M of the universe consists, first of all, of *all* possible events that can occur in the universe. At this stage no time function is defined on M, and we do not assume that one exists *a priori*. However, using certain physical laws together with the cosmological principle, a universal time function t can be *constructed* on M. This assigns a number $t(E)$ to each event, and, by virtue of its construction, the range of t is all *positive* numbers not including zero. Thus, there simply aren't any events such that $t(E) \leq 0$, and, in particular, no event E_{BB} (the big bang itself) such that $t(E_{\mathrm{BB}}) = 0$ exists. Universal time in this sense is similar to absolute temperature as defined in statistical mechanics [see, for example, Buchdahl (1975)]. Here we start with the notion of a system in thermal equilibrium, and then, using certain physical principles, construct a temperature function T that assigns a number $T(S)$ – the temperature of S – to any system S in thermal equilibrium. The function T does not exist *a priori* but must be constructed. The range of the resulting function is all positive numbers not including zero. Thus, systems such that $T(S) \leq 0$ simply do not exist.

One of the most appealing features of the standard model is that it follows logically from Einstein's equation and the cosmological principle. Except possibly for the very early universe, we are on firm ground with Einstein's equation. However, the cosmological principle should be cause for concern. After all, the universe is not exactly isotropic and homogeneous – even on a very large scale – and deviations from isotropy and homogeneity might well imply a nonsingular universe without an initial big bang. That this cannot be the case can be seen from the **singularity theorems** of Hawking and Penrose. These theorems imply that if the universe is approximately isotropic and homogeneous in the present epoch – which is the case – then a singularity must have existed sometime in the past. A very readable account of these singularity theorems can be found in Hawking and Penrose (1996), but for the full details see Hawking and Ellis (1973).

Let us now return briefly to the principle of relativity. We have tacitly assumed that given two events on the world line of an observer (such events are said to have **timelike** separation) there is an absolute sense in which one occurred before the other. For example, I am convinced that my 21st birthday occurred before my 40th birthday. But are we really justified in assuming that "beforeness" in this sense is any more than a type of prejudice common to all human beings and therefore more a part of psychology than fundamental physics? There does, of course, tend to be a very real physical difference between most timelike-separated events. A wine glass in my hand is very different from the same wine glass lying shattered on the floor, and we would be inclined to say that these two events had a very definite and obvious temporal order. However, the physical laws governing the individual glass molecules are completely symmetric with respect to time reversal, and, though highly improbable, it would in principle be possible for the shattered glass to reconstitute itself and jump back into my hand. Of course, such events never happen in practice, at least when one is sober, but this has more to do with improbable boundary conditions than the laws of physics.

For many years it was felt that all laws of physics ought to be time-symmetric in this way. This is certainly true for particles moving under the influence of electromagnetic and gravitational interactions (e.g. glass molecules), but the discovery of weak elementary-particle interactions in the fifties has called into question this attitude. It is now known that there exist physical processes governed by weak interactions (e.g. neutral K-meson decay) that are not time-symmetric. These processes indicate that there does indeed exist a physically objective sense in which the notion of "beforeness" can be assigned to one of two timelike-separated events. Of course, temporal order can be defined with respect to an observer's environment in the universe – if $\rho(E) > \rho(E')$ then, given that the universe is expanding, we would be inclined to say that event E occurred before event E' – but the type of temporal order we are talking about here is with respect to the fundamental laws of physics. A good account of time asymmetry is given in Davies (1974).

Another surprising feature of processes governed by weak interactions is that they can exhibit a definite "handedness." However, unlike that found (on the average) in the human population, which, as far as we know, is a mere accident of evolution, the type of handedness exhibited by weakly interacting processes is universal and an integral part of the objective physical world. It can, in fact, be used to obtain a physical distinction between right-handed and left-handed frames of reference, since an experimental configuration based on a right-handed frame will, in general, yield a different set of measurements from one based on a left-handed frame. For an entertaining discussion of these ideas see Gardner (1967).

Finally, we should say something about the physical units used in this book. Clearly, nature does not care which system of units we use: the time interval between two events on a person's world line is, for example, the same whether she uses seconds or hours as the unit. We shall therefore use a system of units in terms of which the fundamental laws of physics assume their simplest form.

Let us initially agree to use a *second* as our unit of time – we'll choose a more natural unit of time later. We then choose our unit of distance to be a light-second. This is a particularly natural unit of length, since one of the fundamental laws of nature is that light always has the same speed with respect to any observer. By choosing a light-second as our unit of distance we are essentially encoding this law into our system of units. Note that, in units of seconds and light-seconds, the speed of light c is, by definition, unity.

Another feature of light, and one that forms the basis of quantum theory, is that the energy of a single photon is exactly proportional to its frequency. We use this to define our unit of energy as that of a photon with angular frequency one. In terms of this unit of energy, Planck's constant $\hbar$ is, by definition, unity.

Finally, since mass is simply another form of energy (this will be shown when we come to consider special relativity), we also measure mass in terms of angular frequency. For example, if we wish to measure the mass of a particle in units of frequency, we could bring it into contact with its antiparticle. By arranging things such that the resulting explosion consists of just two photons, the frequency of one of these photons will give the mass of our particle in units of frequency.

Since we are defining distance, energy, and mass in units of time, it is important to have a good definition of what we mean by an accurate clock. As a provisional definition, we can define a clock as simply any smoothly running, cyclical device that is unaffected by changes in its immediate environment. This, for example, rules out pendulum clocks. But how can we check that a clock is actually unaffected by changes in its environment? It is no good appealing to some other, better clock – not even the most up-to-date atomic clock, which presumably ticks away the hours in a cellar of the Greenwich observatory – as this would lead to a circular argument. There is only one certain way and that is to appeal to the properties of nature herself. Given a clock, together with the appropriate apparatus, all in a unchanging environment, it is, at least in principle, possible to determine the gravitational constant G in units of time. Recall that we are defining both mass and distance in units of time. If we now change the environment (e.g. by changing the temperature or transferring the laboratory to the moon) and the value of G remains unchanged, we can say, by definition, that we have a good clock and one unaffected by changes in its environment.

The gravitational constant is, of course, defined by $G = ar^2/m$, where a is the acceleration of a particle of mass m caused by the gravitational influence of an identical particle at a distance r. If, for example, we change our unit of time from one second to one minute, then $r \to r/60$ (one light-second $= \frac{1}{60}$ light-minutes), $m \to 60m$ (one cycle per second = 60 cycles per minute), and similarly $a \to 60a$. Thus $G \to G/(60)^2$. A particularly convenient unit of time for gravitational physics is that which makes $G = 1$. This is called a **Planck second** or a **gravitational second**. Whenever we are dealing with gravity we shall use this unit of time; otherwise, for simplicity, we shall stick with ordinary seconds.

Nature gives us many other natural time units. A particle physicist might, for example, prefer to use an electron second (a unit of time such that $m_e = 1$) or even a proton second. It is a remarkable feature of our universe that clocks reading electron, proton, and gravitational time *appear* to remain synchronous.

EXERCISES

1.1 Calculate the following quantities in terms of natural units where $c = \hbar = G = 1$: the mass and radius of the sun, the mass and spin of an electron, and the mass of a proton.

1.2 Another set of natural units is where $c = \hbar = m_e = 1$, where m_e is the mass of an electron. In terms of these units calculate the quantities mentioned in Exercise 1.1.

1.3 You have made email contact with an experimental physicist on the planet Pluto. She wishes to know your age, height, and mass, but has never heard of pounds, feet, or seconds (or any other earthly units). By instructing her to perform a series of experiments, show how this information can be conveyed.

1.4 Your friend on Pluto also wants to know your body temperature. How can this information be conveyed? (Hint: Use your knowledge of statistical mechanics, and choose units such that Boltzmann's constant is unity.)

1.5 According to the principle of relativity there is no preferred state of inertial motion. Does this conflict with the cosmological principle?